

# Best Available Copy

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

**CORRECTED VERSION**

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number  
**WO 02/068666 A1**

- (51) International Patent Classification<sup>7</sup>: C12N 15/82
- (21) International Application Number: PCT/US01/43588
- (22) International Filing Date:  
16 November 2001 (16.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/249,608 17 November 2000 (17.11.2000) US

- (81) Designated States (*national*): AE, AL, AM, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, EE (utility model), EE, ES, FI (utility model), FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KI, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NF, SN, TD, TG).

(71) Applicant (*for all designated States except US*): GENENCOR INTERNATIONAL, INC. [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

Published:  
— with international search report

- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): DUNN-COLEMAN, Nigel [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). LANGDON, Timothy [GB/GB]; Troedyrhiw, Lon Tyllwyd, Llanfarian, Aberystwyth SY23 4UJ (GB). MORRIS, Phillip [GB/GB]; 6 Trefaenor, Comins-coch, Aberystwyth SY23 3UB, Cerdigion, Wales (GB).

(48) Date of publication of this corrected version:  
6 November 2003

(74) Agent: BOYD, Victoria, L.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).

(15) Information about Correction:  
see PCT Gazette No. 45/2003 of 6 November 2003, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 02/068666 A1

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) Abstract: Described herein are methods to enhance the production of more highly fermentable carbohydrates in plants, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from *Aspergillus*, preferably *A. Niger*. The expression vectors may optionally comprise a DNA sequence encoding xylanase from *Trichoderma*, preferably *T. reesei*. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.

5

10 **MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY  
OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES  
ENCODING CELL WALL DEGRADING ENZYMES**

15

**Inventors: Nigel Dunn-Coleman  
Timothy Langdon  
Phillip Morse**

CROSS-REFERENCE TO RELATED APPLICATIONS

20 Pursuant to 35 U.S.C. §119(e), the present application claims benefit of  
and priority to USSN 60/249,608, entitled "MANIPULATION OF THE PHENOLIC  
ACID CONTENT AND DIGESTIBILITY OF FORAGE GRASS CELL WALLS BY  
TARGETED EXPRESSION OF A FERULIC ACID ESTERASE GENE", filed  
November 17, 2000, by Morris et al.

FIELD OF THE INVENTION

25 This invention relates to methods to enhance to availability of fermentable  
carbohydrates.

BACKGROUND OF THE INVENTION

30 The present crisis in livestock agriculture has prompted a resurgence of  
interest in grass-fed animals. However, while a high-forage diet may be desirable, it  
does not currently satisfy the demands of modern animal production. For the animal  
to make efficient use of the forage it consumes, the energy demands of the micro-  
organisms in the rumen must be met and synchronized with the availability of plant  
proteins. Otherwise this lack of synchrony will lead to (a) proteins and other  
35 nutrients being poorly utilized in the rumen, (b) loss of nitrogen, in urine and feces  
and therefore, the environment and (c) the need to feed excessive amounts of  
protein concentrates as supplements to the ruminant diet.

Cellulose and hemicellulose in grass and maize tissues could meet the

energy requirements of the ruminant or provide new feed-stocks for industrial fermentation to ethanol. This potential is not currently realized because the cell walls are lignified and the cell wall polysaccharides highly cross-linked with phenolic residues and lignin, resulting in low rates of plant cell wall digestion in comparison to rates of protein breakdown in ruminants. This is a particular problem for the most important forages in Europe, the ryegrasses *Lolium perenne* and *L. multiflorum* as well as one of the major impediments to the wider use of better adapted species, such as *Festuca arundinacea*, as a forage crop. Increasing the digestibility index of grasses has therefore been a major breeding objective for several decades but progress has been slow due to difficulties in fixing natural variation in the synthetic varieties derived from these outbreeding species (Hayward, *et al.*, TAG 70:48 (1985)).

Removing labile phenolics by chemical treatment with alkali is known to increase the biodegradability and nutritional value of low-quality feed such as cereal straw, and is employed commercially for feed upgrading. Reducing phenolic cross-linking of cell wall carbohydrates is therefore a predictable way of improving the rate of digestion and digestibility of ryegrass. However chemical modification may have other disadvantages. Therefore, genetic modification would be a preferable method of changing the cell wall chemistry of highly digestible varieties. Many in the field are pursuing this approach. An alternative, however, is to use genetic modification to reduce the levels of phenolic acids in the cell walls available for crosslinking either by directly disrupting ester bonds linking phenolics and lignins to cell wall polysaccharides or by preventing excessive ferulation of cell wall carbohydrates prior to their incorporation into the cell wall.

This invention meets this and other needs by using targeted or inducible expression of cell wall degrading enzymes in plants.

### SUMMARY OF THE INVENTION

Provided herein are methods for enhancing the availability of fermentable carbohydrates. In one aspect, there is provided an expression cassette  
5 comprising a DNA sequence encoding at least one cell wall degrading enzyme. The DNA sequence encoding at least one cell wall degrading enzyme may be operatively linked to a promoter sequence. The promoter may be constitutive or inducible. The expression cassette may further comprise a targeting sequence.

In one embodiment, the cell wall degrading enzyme is selected from the  
10 group consisting of ferulic acid esterase, xylanase, xylosidase, cellulase, endoglucanase, and cellbiohydrolase. In a preferred embodiment cell wall degrading enzyme is derived from a fungal source. In a more preferred embodiment, the fungal ferulic acid esterase is an *Aspergillus* ferulic acid esterase, preferably *A. niger*. In another embodiment the xylanase is derived  
15 from *Trichoderma*, preferably *T. reesei*.

In another aspect of the invention, there is provided a plant transformed with the expression cassette comprising a DNA sequence encoding at least one cell degrading enzyme. The plant may be selected from the group consisting of  
20 Festuca, Lolium, Avena and Zea. In a preferred embodiment the plant is a forage grass. In another embodiment, the plant is maize.

Further provided herein is a method of controlling the level of phenolic acids in plant cell walls of a transgenic plant. The method, in one embodiment, comprises introducing to a plant cell an expression cassette comprising a DNA  
25 sequence encoding at least one cell wall degrading enzyme, preferably a ferulic acid esterase.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however,  
30 that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.



### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a restriction map of a DNA fragment containing the gene encoding the 38kd ferulic acid esterase.

5 Figures 2 A-E illustrate the complete DNA (SEQ. ID NO:\_\_\_), with highlighting to point out the signal sequence, intron and various restriction endonuclease sites, and amino acid sequence (SEQ. ID. NO:\_\_\_) corresponding to the gene encoding the 38 kD ferulic acid esterase isolated from *Aspergillus niger*.

10 Figure 3 illustrates the DNA sequence of the gene encoding the 38 kD esterase (SEQ. ID. NO:\_\_\_).

Figure 4 illustrates the construction of the intronless ferulic acid esterase isolated from *Aspergillus niger*.

15 Figure 5 illustrates that the overlapping of PCR products made with primers FAE-I5 FAE-I3 creates two possible uninterrupted reading frames -- the top in the figure below is functional (highlighted serine is at active site), the bottom is inactivated.

Figure 6 illustrates the possible vector constructions useful in the present invention. Various combinations are possible. Although and FAE gene is depicted another cell wall degrading enzyme may be used alone (i.e., instead of) or in conjunction with the FAE gene. Amp = ampicillin resistance gene.

20 Figure 7 illustrates pCOR105.

Figure 8 illustrates a generic ALE-TER vector.

Figure 9 illustrates the KDEL-COOH ER retention sequences.

Figure 10 illustrates the FAE-LINKER-FRAMESHIFT structure and sequence.

25 Figure 11 illustrates plant transformation cassettes.

Figure 12 is a table of the vectors used herein.

Figure 13 depicts the barley aleurain vacuolar and apoplast signal sequence.

Figure 14 illustrates the rat sialyl transferase structure and sequence.

30 Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure and sequence.

Figure 16 illustrates the targeted expression of gfp to different cell compartment. Also shown are schematics of the vectors used.

Figure 17 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under ER and APO targeting sequences.

5 Figure 18 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under Vac targeting sequence.

Figure 19 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves of different ages.

10 Figure 20 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves under Vac, ER and APO targeting sequences.

Figure 21 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under Vac targeting sequence.

15 Figure 22 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under APO and ER targeting sequence.

Figure 23 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Festuca arundinacea* plants expressing FAE under an actin promoter.

20 Figure 24 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Lolium mutflorum* plants expressing FAE under an actin promoter.

Figure 25 illustrates the rate of fermentation and cumulative gas production in *Festuca arundinacea* cells.

Figure 26 illustrates the *in vitro* fermentation of *Festuca arundinacea* cell walls from cell cultures expressing recombinant FAE1.

25 Figure 27 illustrates the Time to maximum rate digestion for *Festuca arundinacea* cells.

Figure 28 illustrates the total gas production in *Festuca arundinacea* cells.

Figure 29 illustrates the kinetics of FAE activity by ferulic acid release from cell wall under self digestion in *Festuca arundinacea* and stimulation by xylanase.

30 Figure 30 illustrates the beta-glucuronidase activity under the *Lolium* See1 senescence promoter in leaves of transgenic plants of *Lolium mutflorum*.

Figure 31 illustrates the release of monomeric and dimeric HCAs on self digestion of leaves of vacuolar targeted FAE expressing plants.

Figure 32 is a schematic of the pTP10-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

5 Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector.

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

10 Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector.

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector.

15 Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector.

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

20 Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector.

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector.

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector.

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector.

30 Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector.

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector.

Figure 47 is a schematic of the pUG4 vector. Also shown is the 5295 bp nucleotide sequence of the vector.

Figure 48 is a schematic of the pUB8.11 vector. Also shown is the 5001 bp nucleotide sequence of the vector.

5 Figure 49 is a schematic of the pTP11-1 vector. Also shown is the 5387 bp nucleotide sequence of the vector.

Figure 50 illustrates the actin promoter and its corresponding nucleotide sequence.

10 Figure 51 illustrates the Aleurain-NPIR delete structure. The corresponding nucleotide sequences are also shown.

Figure 52 illustrates the SEE1 (senescence enhanced) promoter sequence.

Figure 53 illustrates the SEE1 (senescence enhanced) promoter sequence plus the vacuolar aleurain signal/NPIR sequence.

15

#### DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred  
20 to herein are expressly incorporated by reference.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D Ed., John Wiley and Sons, New York  
25 (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are  
30 described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy

orientation, respectively. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

5           The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

#### Definitions

10           It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense  
15 including "and/or" unless the content clearly dictates otherwise.

          "Conservatively modified variants" applies to both amino acid sequences and polynucleotides. With respect to particular polynucleotides, conservatively modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not  
20 encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical polynucleotides encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the  
25 corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every polynucleotide herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG,  
30 which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in each described

sequence. For purposes of protein expression, there are "sub-optimal codons." These are codons that are not preferred by a particular genus or species. Altering these "sub-optimal codons" to "preferred codons" is a silent mutation in that the amino acid encoded by the codons is the same but one codon is  
5 preferentially expressed by the particular genus, *e.g.*, *Triticum spp.*

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a polynucleotide, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively  
10 modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 15 1) Alanine (A), Serine (S), Threonine (T);
  - 2) Aspartic acid (D), Glutamic acid (E);
  - 3) Asparagine (N), Glutamine (Q);
  - 4) Arginine (R), Lysine (K);
  - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
  - 20 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (*see, e.g.*, Creighton, *Proteins* (1984)).

"Pyroglutamic acid" is the cyclized internal amide of L-glutamic acid

The phrase "controlling the level of phenolic acids" refers to the manipulation of phenolic acid expression in plants, particularly plant cell walls.  
25 The manipulation can be either positive; *e.g.*, increasing the levels of phenolic acids; negative, *e.g.*, decreasing the level of phenolic acids; or neutral, *e.g.*, changing the relative amounts of specific phenolic acids in the cell walls but keeping the total amount relatively the same. The timing of manipulation can be during plant growth or after plant growth, *e.g.*, after a plant has been cut or pulled  
30 from the ground or ingested. "Plant cell walls" refers to the cell walls of any cell of the plant.

The term "derived" means that a polynucleotide or protein is related to

another polynucleotide or protein. The relations can be one of homology, *e.g.*, nucleotides and proteins from certain species are homologous to similar polynucleotides and proteins of other species; analogy, *e.g.*, proteins perform the same function and therefore are related to each other regardless of organism of origin. The relationship can be a man-made one, *e.g.*, a protein (and a polynucleotide) can be derived from another protein by mutation; or chemical manipulation (peptidomimetics). Furthermore, a protein or a polynucleotide can be derived from an organism if, in the natural state, the protein or polynucleotide is found in one organism but recombinantly produced in another.

10 The term "exogenous polynucleotide" refers to a polynucleotide which is introduced into the plant by any means other than a sexual cross or sexual reproduction. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous polynucleotide is referred to here as an R<sub>1</sub> generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are progeny of such a plant.

The term "isolated polynucleotide molecule" or "isolated protein" refers to a polynucleotide or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *FAE1* gene is separated from open reading frames which flank the gene and encode a protein other than *FAE1*.

A "*FAE1* encoding polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of an *FAE 1* gene or which encodes a *FAE1* polypeptide. *FAE1* polynucleotides can also be identified by their ability to hybridize under low stringency conditions (see below) to nucleic acid probes having a sequence of 8 to 300 bases, preferably a sequence of 80 to 100 bases

in the sequence disclosed in WO 98/14594.

The term "nucleic acid encoding," "nucleic acid sequence encoding" or "polynucleotide encoding" refers to a polynucleotide which directs the expression of a specific protein or peptide. The polynucleotides include both the DNA strand  
5 sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The polynucleotides include both full length polynucleotides as well as shorter sequences derived from the full length sequences. It is understood that a particular polynucleotide includes the degenerate codons of the native  
10 sequence or sequences which may be introduced to provide codon preference in a specific host cell. The polynucleotide includes both the sense and antisense strands as either individual single strands or in the duplex form.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

15 The term "plasmid" refers to a circular double stranded DNA molecule which comprises the coding sequence of interest, regulatory elements, a selection marker and optionally an amplification marker. A plasmid can transform prokaryotic cells or transfect eukaryotic cells. An "expression cassette" means a portion of a plasmid (or the entire plasmid) containing the regulatory  
20 elements desired for transcription, translation and/or expression and the coding region of a polynucleotide. A plasmid may contain one or more expression cassettes. If multiple expression cassettes are introduced into a plant, they can be introduced simultaneously or at different times. If simultaneous introduction is desired, the expression cassettes can be on one plasmid or more. Typically, an  
25 expression cassette comprises a promoter, poly A+ tail, and signal sequences that target the expressed polypeptide to a specific region of a cell or to be secreted, if desired. Examples of signal sequences that "target expression" of ferulic acid esterase include sequences located upstream of the FAE coding sequence. The polynucleotide that encodes the signal sequence is found  
30 preferably within the 100 nucleotides "upstream" (in the 5' direction) from the initiation codon (AUG). More preferably, the polynucleotide that encodes the signal sequence is found within the 50 nucleotides upstream from the initiation



codon. Many different cellular organelles are targeted by the signal sequences used in this invention. The organelles include, but are not limited to, vacuoles, Golgi apparatus, endoplasmic reticula, and apoplasts. In addition to upstream signal sequences, the expression cassette of this invention may include a  
5 polynucleotide that encodes a signal sequence at the 3' end. These signal sequences include, but are not limited to stop codons and the KDEL sequence. In addition to KDEL, other similar sequences are contemplated by this invention, including but not limited to RDEL. In addition to a KDEL sequence, a signal sequence can include a linker to a KDEL sequence. A linker is an extension of  
10 the reading frame of the encoding polynucleotide to the signal sequence. Preferably, the polynucleotide encoding the signal sequence is directly downstream from the coding sequence, more preferably less than 100 base pairs from the stop codon, more preferably less than 20 base pairs from the stop codon.

15 The term "polynucleotide," "polynucleotide" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses polynucleotides containing known analogues of natural nucleotides which have similar binding properties as the reference polynucleotide  
20 and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular *FAE1* polynucleotide of this invention also implicitly encompasses conservatively modified variants thereof (*e.g.* degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions  
25 may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.*, 1992; Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term polynucleotide is used interchangeably with  
30 gene, cDNA, and mRNA encoded by a gene.

The term "polypeptide," "peptide," and "protein" are used interchangeably and refer to amino acids connected by peptide bonds. Polypeptides can be

entire proteins or portions thereof. For Example, a FAE1 polypeptide may refer to the entire FAE1 protein or fragments of the FAE1 protein. A "ferulic acid esterase with an altered glycosylation site" refers to a FAE protein wherein a mutation has changed the glycosylation pattern of the protein. Mutations that effect such changes are well known in the art and include, but are not limited to, amino acid substitutions, and mutations in the proteins of the Golgi apparatus and endoplasmic reticulum that effect glycosylation of proteins.

The term "promoter" refers to a polynucleotide that directs expression of a coding sequence. A promoter can be constitutive, *i.e.*, relatively independent of the stage of differentiation of the cell in which it is contained or it can be inducible, *i.e.*, induced by specific environmental factors, such as the length of the day, the temperature, *etc.* or a promoter can be tissue-specific, *i.e.*, directing the expression of the coding sequence in cells of a certain tissue type. A "senescence" promoter is an inducible promoter that causes transcription to be initiated upon a certain event relating to age of the organism. A "heat shock promoter" is an inducible promoter that causes transcription to be initiated upon a change in temperature. An example of a heat shock protein promoter is the Soybean Gmhsp promoter. In addition to these inducible promoters, one of skill will realize that other inducible promoters can be used. For example, a wound induced promoter, like LAP. See, US Patent No. 5,962,670.

The term "purified" denotes that a polynucleotide or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polynucleotide or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "specifically hybridizes" refers to a nucleic acid probe that hybridizes, duplexes or binds to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and

annealing conditions, see, for example, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) ("Sambrook") or CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

5           The term "stringent conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations refers to sequence dependent, binding and washing environments. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) LABORATORY  
10           TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY—HYBRIDIZATION WITH NUCLEIC ACID PROBES part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and  
15           pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the  $T_m$  for a particular probe. An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin  
20           with 1 mg of heparin at between 40 and 50°C, preferably 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at from 70 to 80°C with 72°C being preferable for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at about 60 to 70°C, preferably 65°C for 15 minutes (see, Sambrook, *supra* for a  
25           description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 40 to 50°C, preferably 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 35 to 45°C, with 40°C  
30           being preferable, for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Polynucleotides which do

not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a polynucleotide is created using the maximum codon degeneracy permitted by the genetic code.

5           The term "transgenic plant" refers to a plant into which exogenous polynucleotides have been introduced and their progeny. Typically, cells of a plant are transformed with the exogenous polynucleotide and a transgenic plant is regenerated from the transformed cells. The regenerated plant is then bred to produce a strain of transgenic plants.

10           "Xylanase" (EC 3.2.1.8) refers to a well described class of glycosyl hydrolases that hydrolyze xylan. Commercial applications of xylanase include the degradation and bleaching of wood pulp for paper making. Xylanase can also be added to animal feed to improve the digestibility of plant matter. Typically, commercial xylanase is derived from fungi. A preferred xylanase is derived from  
15 *Trichoderma*.

#### Preferred Embodiments

Plant cell walls contain a range of alkali-labile ester-linked phenolic acids. In particular, grass cell walls are characterized by the presence of large amounts of esterified ferulic and p-coumaric acids (mainly in their E configurations), linked  
20 to arabinoxylans at the C5 of arabinose. These are released as ferulated oligosaccharides (FAX and PAX) by cellulase treatment but *in vivo* provide a substrate for peroxidase-catalyzed cross-linking of cell wall polysaccharides and lignin. The high levels of these phenolic acids and their dimers have a dramatic influence on the mechanical properties, digestibility and rates of digestion of  
25 grasses by ruminants.

Previous work has shown that ferulic acid is the predominant p-hydroxycinnamic acid esterified to grass polysaccharide but until recently the only  
ferulic acid dehydrodimer to have been isolated was 5,5'-diferulic acid. Recently  
30 new dehydrodiferulate dimers and cyclobutane-type dimer mixtures have been isolated from plant cell walls (Waldron, *et al.*, *Phytochemical Analysis* 7:305 (1996)). As can be seen in Figure 1, these mixtures are present in large amounts in grass cells. Ether linked ferulic acid-coniferyl alcohol dimers, have also been isolated from cell walls (Jacquet, *et al.*, *Polyphenol Comm. Bordeaux*

pp451 (1996)) establishing for the first time that ferulate esters are oxidatively co-polymerized with lignin precursors which may anchor lignins to cell wall polysaccharides. The yield of these dimers in grass cells indicates that phenolic dehydrodimer cross-linking of cell wall polysaccharides is much more extensive than was previously thought.

An enzyme system has been reported from parsley endomembranes that catalyses the ferulation of endogenous polysaccharide acceptors from feruloyl CoA, pointing to the ER/golgi as the site of polysaccharide esterification and the CoA ester as the physiological co-substrate (Meyer, *et al.*, *FEBS Lett.* 290:209 (1991)). Further evidence for this has been found in water-soluble extracellular polysaccharides excreted in large amounts into the medium by grass cell cultures. This material is highly esterified with ferulic and p-coumaric acid at levels similar to the cell walls of the cultured cells.

Feruloyl esterase activity has been detected in several fungal species including, anaerobic gut fungi, yeasts, actinomycetes, and a few fiber-degrading ruminal bacteria, which enables them to de-esterify arabinoxylans and pectins.

Two ferulic acid esterases (FAE), distinguished on the basis of molecular weight and substrate specificity, have been isolated from *Aspergillus niger* and have been shown to quantitatively hydrolyze ferulic acid and release dehydrodiferulate dimers from plant cell walls. Furthermore, FAE has been observed to act synergistically with xylanase to release ferulic acid from plant cell walls at a higher rate. Recently, a ferulic acid esterase (FAE) gene has been cloned from *Aspergillus niger* (Michelson, *et al.* European Patent Application No. 9510370.1). The inventors have found the recombinant enzyme releases ferulic acid and diferulate dimers from grass cell walls in a concentration dependent manner and that this enzyme is stable at 30°C pH 5.0 in the presence of substrate and has a half life of 61 h at 30°C in the presence of vacuolar extracts (pH 4.6) of grass cells. This gene was, therefore, a candidate for targeted and inducible expression of FAE in grasses (*e.g.*, *Lolium multiflorum*).

The present invention provides for methods of changing the cell wall structure of transgenic plants and therefore, making them more digestible. The method comprises introducing a ferulic acid esterase coding sequence into the

cells of a plant. Operably linked to the coding sequence is a promoter that can be either constitutive or inducible and signal sequences that serve to target expression of the coding sequence in the desired organelle in the desired cell of the plant. The signal sequences can be either or both N terminal or C terminal sequences.

Optionally, a second and/or third coding sequence is introduced into the plant. It is preferred that a fungal xylanase coding sequence be coexpressed with the FAE coding sequence.

This invention also provides for transgenic plants which contain FAE1 coding sequences, leading to more digestible grasses.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. Basic texts disclosing the general methods of use in this invention include Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); and Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

#### A. Isolation of Polynucleotides

The isolation of the polynucleotides, *e.g.*, FAE1 and xylanase of the invention may be accomplished by a number of techniques. See, for example, copending US application 08/952,445 which describes the isolation of a FAE from *Aspergillus niger*, and copending US application 09/658,772 which describes the isolation of a xylanase from *T. reesei*.

For instance, oligonucleotide probes based on the sequences cited here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, *e.g.*, using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be

packaged into the appropriate vector. To prepare a library of cDNA from a specific cell culture, e.g., *Aspergillus niger*, mRNA is isolated from the culture and a cDNA library containing the gene transcripts is prepared from the mRNA.

5 The cDNA or genomic library can then be screened using a probe based upon the sequence of a known polynucleotide such as the polynucleotides cited here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. In addition to probes derived from known polynucleotides, degenerate probes may be used. Techniques for making and using degenerate probes are well known in the art and can be found in Sambrook and Ausubel.

10 Alternatively, the polynucleotides of interest can be amplified from polynucleotide samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone polynucleotides that code for proteins to be expressed, to make polynucleotides to use as probes for detecting the presence of the desired mRNA in samples, for polynucleotide sequencing, or for other purposes.

20 Appropriate primers and probes for identifying ferullic acid esterase-specific genes, as well as xylanase sequences, from fungi and plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Reaction components are typically: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200  $\mu$ M dATP, 200  $\mu$ M dCTP, 200  $\mu$ M dGTP, 200  $\mu$ M dTTP, 0.4  $\mu$ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.

30 Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, *et al.*, *J. Am. Chem.*

Soc. 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

5            Suitable sources for the ferulic acid esterase used in this invention include but are not limited to, *Neurospora crassa*, *Aspergillus spp.* and specifically, *Aspergillus niger*. The xylanase used in this invention can be derived from any suitable source including, but not limited to, *Trichoderma reesei* and *Aspergillus spp.*

10

### B. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of plant species are well known and described in the technical and scientific literature. See, for example, Weising, *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length FAE1 protein, will preferably be combined with transcriptional and translational initiation and targeting regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant under the desired conditions.

Promoters can be identified by analyzing the 5' sequences of a desired gene. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

25  
30



A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell* 1:855-866 (1989); Bustos, *et al.*, *Plant Cell* 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7:4035-4044 (1988); Meler, *et al.*, *Plant Cell* 3:309-316  
5 (1991); and Zhang, *et al.*, *Plant Physiology* 110:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions  
10 and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the actin and ubiquitin promoters and other transcription initiation regions from various plant genes known to those of skill. A particularly preferred constitutive  
15 promoter is the rice actin promoter (see, McElroy, *Plant Cell*, 2:163 (1990)).

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control  
(inducible promoters). Examples of tissue-specific promoters under  
20 developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots or seeds.

In one aspect of the instant invention, expression of FAE occurs after the the plant has been cut, removed from the ground or ingested. Thus an appropriate promoter would be a senescence promoter. For example, *BFN1* has  
25 recently been shown to be a nuclease expressed in senescing leaves, Perez-Amador, *et al.*, *Plant Physiol.* 122:169 (2000). Similarly, SAG12, a cysteine protease is also found in senescing leaves (Noh & Amasino, *Plant Mol. Biol.* 41:181 (1999). In a preferred embodiment, the promoter from the *gem* gene of *Festuca pratensis* is used to direct expression of FAE in senescing leaves.

30 In another aspect, the FAE would be expressed upon ingestion by a foraging animal. Exemplary promoters for this aspect would include Soybean Gmhsp 17.5 promoter and the leucine aminopeptidase (LAP) promoter. The

GMhsp promoter is from a heat shock protein gene and initiates expression if the temperature of the environment is increased. In the laboratory, an increase of 15°C for 2 hours is the preferred heat shock. However, in non-laboratory conditions suitable increases in temperature will occur in silos and in the rumen of animals that have ingested the plants of this invention. The LAP promoter initiates the expression of the FAE gene upon wounding of the plant. Such wounding would occur after cutting the plant or after mastication by a foraging animal. Tissue specific promoters that could be used in this invention include promoters of genes that are differentially expressed in the leaves of grasses. An example of a leaf specific promoter is the *rbcS* promoter of tomato (*Proc. Nat'l Acad. Sci. USA* 84:7104 (1987)). This promoter normally regulates a gene determined to be important in photosynthesis.

For proper polypeptide expression, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural fungal gene, from a variety of other fungal or plant genes, or from T-DNA. These sequences are well known and readily available to those of skill in the art.

In addition to a promoter and poly A+ sequences, the preferred expression vectors of this invention also will contain signal sequences. These are polynucleotides found at the 5' and/or 3' ends of the coding region and serve to target expression of the gene to specific cellular organelles. These signal sequences can be both upstream or downstream of the coding region. Some preferred examples of upstream signal sequences include the barley aleurain sequence (Rogers, *Proc. Nat'l Acad. Sci. USA* 82:6512 (1985) which targets vacuoles and the *Aspergillus* apoplast signal. This signal sequence targets expression to the apoplast.

In addition to targeting expression to specific organelles, it may be desirable to retain the expressed FAE in the Golgi or endoplasmic reticulum. The well known ER retention signal, KDEL, can be added to the 3' end of the coding polynucleotide.

The vector comprising the expression cassettes (e.g., promoters and/or coding regions) of the invention will typically comprise a marker gene which

confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to hygromycin, kanamycin, G418, bleomycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

5

### C. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment or the constructs may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

See Dalton et al. (Co-transformed, diploid *Lolium perenne* (Perennial Ryegrass), *Lolium multiflorum* (Italian Ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. *Plant Cell Reports* (1999) 18(9), 721-726) for exemplary methods for culturing and transformation of grasses.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, et al., *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, et al., *Proc. Natl. Acad. Sci. USA* 82:5824 (1985).

*Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, et al., *Science* 233:496-498 (1984), and Fraley, et al., *Proc. Natl. Acad. Sci. USA* 80:4803 (1983). US Patent 5,591,616 discloses *Agrobacterium* mediated transformation techniques in monocotyledons.

Ballistic transformation techniques are described in Klein, et al., *Nature*

327:70-73 (1987). In a preferred embodiment, a particle in-flow gun (PIG) is used to transform the plant cells of this invention.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that  
5 possesses the transformed genotype and thus the desired phenotype such as improved digestibility. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is  
10 described in Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillian Publishing Company, New York, 1983; and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are  
15 described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

To determine the presence of or increase of FAE1 activity, an enzymatic assay can be used or an assay to measure increases and decreases in rates of fermentation. These assays are readily available in the literature and those of skill in the art can readily find them.

20 One of skill will recognize that other assays can be used to detect the presence or absence of FAE1. These assays include but are not limited to; immunoassays and electrophoretic detection assays (either with staining or western blotting).

The polynucleotides of the invention can be used to confer desired traits  
25 on essentially any plant. However, the main utility of this invention is in the improved digestibility of forage plants. Thus, it is envisioned the transgenic plants of this invention will include but not be limited to the following genera: *Lolium*, *Festuca*, *Triticum*, *Avena*, and *Medicago*. The FAE1 genes of the invention are particularly useful in the production of transgenic plants in the  
30 genus *Lolium*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be

introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

As mentioned above, the transgenic plants of this invention can be used as a foraging crop for animals, such as cattle, sheep, goats and horses. In addition, the methods of this invention can be used to transform any plant into which FAE expression is desired. For example, it is advantageous to break down cell walls during biomass conversion or during processing of plants for foodstuffs. This invention would help to achieve this goal more effectively and inexpensively.

10

The inventive methods herein may also be used to provide additional enzymes to enhance the availability of fermentable sugars in plants. Plant carbohydrates may be subject to further modification, either exogenously or endogenously, by the action of other enzymes. Such enzymes include, but are not limited to, endoglucanases, xylosidases and/or cellobiohydrolases. These enzymes may be provided either in an expression cassette provided for herein (i.e., endogenous) or applied to the plant cell walls (i.e., exogenous) to enhance the availability of mono- and/or di-saccharides.

Plants other than grasses may find a use in the present invention. For example, corn (or maize) is specifically contemplated to be useful. The grass Festuca is similar to maize in cell wall structure and therefore provides a good model of the ability to enhance fermentable carbohydrates in corn. Other useful plants contemplated for use in the present invention are Festuca, Lolium, Zea, Avena, Sorghum, Millet (tropical cereals), Miscanthus (a grass with potential for use as a biomass energy crop), Cenchrus, Dichanthium, Bracharia and Paspalum (apomictic tropical range grasses) and Poa (Kentucky bluegrass).

Cell walls of forage grasses makes up 30-80% of forage dry matter representing a major source of energy for ruminants, but less than 50% of this fraction is digested by the animal. Conversion of low-value biomass to sugars and ethanol is also less than optimal due to the carbohydrate unavailability of the

feedstocks, including but not limited to bagasse, rice straw, corn stover and corn fiber.

Ferulic and other hydroxycinnamic acids are ester linked to arabinosyl residues in arabinoxylans, and play a key role in crosslinking xylans to lignin, resulting in less degradable cell walls. Ferulic acid esterase (FAE) can release both monomeric and dimeric ferulic acid (FA) from arabinoxylans making the cell wall more susceptible to further enzymatic attack. Transgenic plants have been produced expressing an FAE gene following microprojectile bombardment of cell cultures. Measurements of the level of FAE activity from different vectors targeting FAE to the vacuole, ER and apoplast under constitutive or inducible (heat shock) promoters shows that at least for constitutive expression of vacuolar targeted FAE, the activity was highest in young leaves and increased along the leaf lamina. We also show that FAE expression results in release of monomeric and dimeric FA from cell walls on cell death and this was enhanced several fold by the addition of xylanase. An effect of FAE expression on the monomeric and dimeric cell wall ester linked ferulate content in comparison to control (non-transformed) plants is seen. Generally, the lower the levels of monomers and, in particular, dimers of hydroxycinnamic acids in leaves, the higher the digestibility and/or availability of complex carbohydrates for conversion.

Senescence is the terminal phase in leaf development and occurs without growth or morphogenesis. Therefore the metabolism/physiology of this stage of the leaf's lifespan can be targeted directly for alteration with minimal detrimental impact on early development. Senescence follows leaf maturity and is associated with the expression of specific genes. These genes and their controlling elements can be exploited to manipulate development, adaptation, productivity and quality traits in crop plants. There seems to be good conservation of senescence physiology across the range of higher plant species and thus these promoters are useful in the present invention.

The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but

merely as being illustrative and representative thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar);  $\mu$ M (micromolar); N (Normal); mol (moles); mmol (millimoles);  $\mu$ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms);  $\mu$ g (micrograms); L (liters); ml (milliliters);  $\mu$ l (microliters); cm (centimeters); mm (millimeters);  $\mu$ m (micrometers); nm (nanometers); ° C. (degrees Centigrade); h (hours); min (minutes); sec (seconds); msec (milliseconds); Ci (Curies) mCi (milliCuries);  $\mu$ Ci (microCuries); TLC (thin layer chromatography); Et (ethyl), Me (methyl).

10

### Example 1 Preparation of Enzyme Encoding DNA Sequences

A genomic clone for FAE1 (see Figures 1-3) was used as the starting point for the preparation of an intronless FAE1 encoding DNA sequence. The sequence for the genomic clone is given in Figures 2 and 3. Separate fragments for both FAE exons were recovered by PCR from a 5.5kb EcoRI fragment of the genomic clone in pLITMUS28, and 'cDNA' created by overlapping PCR. See Figure 4.

Two 5' primers were used. FAE-S5 which amplifies the entire reading frame (including the Aspergillus signal), and FAE-N5 which amplifies only the mature protein (i.e. has no signal). A number of codons are optimised (underlined in primer sequences below). The overlap product may be derived from either FAE-I5 (wild type) or FAE-I3 (conserved Ser changed to Ala) primers, allowing production of enzymatically inactive protein to check toxicity. As shown in Figure 5, overlapping of PCR products made with FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames. If the complement to FAE-I5 serves as the template when recombined then the encoded protein retains the serine moiety and the esterase is functional (highlighted serine is at active site). If the FAE-I3 primer serves as the template the serine is replaced with an alanine and the esterase is inactivated (highlighted alanine in bottom amino acid sequence given in Figure 5).

25  
30





(Pst1) (Not 1)

(AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAG  
 AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT  
 5 GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT  
 AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTA  
 GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC  
 AAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA  
 AGCTT CTA GATCT (where K=G or T)  
 10 (HindIII) (XbaI)

A redundancy in the TER5 primer (GCGKAA) creates fragments having either a stop codon (TAA) or glutamate codon (GAA) in one reading frame. The glutamate codon is in frame with a downstream KDEL motif.

15 The fragment and modified pCOR105 vector were cut with PstI and XbaI, according to manufacturers instructions, relevant fragments gel-purified, ligated with T4 DNA ligase and transformed into *E. coli*. Resulting clones were then sequenced to establish which TER5 alternatives were present.

Initial FAE expression vectors were then constructed from these vectors  
 20 by inserting FAE-S5/FAE-3 PCR products (T4 DNA polymerase 'polished' in the presence of dNTPs, purified and digested with NotI, cloned into EcoRV and NotI digested vector) or FAE-N5/FAE-3 PCR products (purified and NotI digested, cloned into NotI digested and calf intestinal alkaline phosphatase treated vector).

The initial pCOR105-*nos* terminator clones were also modified by the  
 25 addition of ALE-5/ALE-3 PCR products (encoding wild-type and modified barley aleurain signal peptides, see below for details). The products were 'polished' with T4 DNA polymerase in the presence of dNTPs, purified and cut with NotI, then cloned into EcoRV and NotI digested vectors. Addition of the ALE sequences creates a series of vectors which can express a reading frame inserted at the  
 30 NotI or NcoI sites as a fusion to the barley aleurain signal, with or without vacuolar targeting motif, and with or without an ER retention motif. HindIII sites flanking the translation initiation codon and transcriptional terminator allow easy

movement of transcription units between expression vectors providing different promoter sequences. (See Figure 8 depicting the generic ALE-TER vector.)

Vector sequences were confirmed by sequencing. Two artifacts were found. Firstly, the redundant codon in TER5 was found to be AAA in one clone, which was subsequently used as the source of all KDEL fusions (the peptide sequence is KPLKDEL, rather than EPLKDEL as designed). See Figure 9. Secondly, an additional base is found at the site of the redundant codon in one clone, creating a frameshifted terminal peptide (ETTEG, Figure 10) which was used as a control in some constructs.

Exploitation of the modular arrangement of signal peptides in the above vector series allowed various combinations of FAE and targeting motifs to be created using standard molecular biology procedures (i.e., restriction digest, purification of relevant fragments and ligation as appropriate). For example, the NotI fragment containing the FAE reading frame was inserted into the NotI site of the frameshifted clone described above to create vector **pTP3.1**. The native Aspergillus COOH-terminus was inserted into a FAE-S5/FAE-3 clone as a SphI (T4 DNA polymerase polished) – NcoI fragment from the FAE genomic clone (replacing the NotI (T4 DNA polymerase polished) – NcoI fragment), creating vector **pTP4a2**, which then encodes the entire, unmodified, Aspergillus FAE. Replacement of the Sall/XbaI fragment of pTP3.1 with that of pTP4a2 then created **pTP11.1**, which encodes FAE with a native Aspergillus COOH-terminus but a barley aleurain N-terminal signal.

Briefly, other vectors made in this series were; **pTP8.5**, the FAE NotI fragment inserted into the NotI site of an ALE-frameshifted COOH-terminus construct, aleurain N-terminus; **pTP5.1**, replacement of the native Aspergillus COOH terminus with a KDEL peptide (NotI/XbaI fragment exchange), Aspergillus N-terminal signal retained; **pTU4.4**, BamHI fragment of pTP11.1 replaces BamHI fragment of pTP5.1, creates FAE reading frame fused to heterologous N- and C-termini (aleurain signal and KDEL).

Vectors in which the aleurain vacuolar targeting motif NPIR was replaced by NPGR (found to be inactive in some plant assays) were created by replacing an EcoRV/NotI fragment with ALE PCR product which had been cut with AccI

- (T4 DNA polymerase polished) and NotI (vectors pTT5.5 and pTT5.14, Aspergillus COOH-terminus). The BamHI fragment of pTT5.5 was used to replace that of pTP5.1 to produce pTU5, creating an FAE reading frame fused to heterologous N- and C-termini (NPGR modification of aleurain signal and KDEL).
- 5 The aleurain signal was also modified by PCR mutagenesis to remove the vacuolar targeting NPIR motif in its entirety (directed by primer ALECUT, which contains a NotI site to allow exchange of BglII/NotI fragments). NPIR deletion was created in this way in pTP11.1 (creating pUA4.4), and in pTP5.1 by exchange of BamHI fragments with pUA4.4 (creating pUG4).
- 10 Finally, PCR mutagenesis, using overlap of fragments generated by primers GLY3 and GLYB, was also used to alter a potential glycosylation site (asparagine codon changed to aspartate, as carried out for example in Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in Aspergillus awamori glucoamylase by site-directed
- 15 mutagenesis to eliminate N-glycosylation and inactivation by deamidation; see sequence data for exact change, vector pTP10.1).

PCR primers

- 20 TER-5 (SEQ ID NO:     )  
 AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG  
 AAGCAGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is  
 underlined.)
- TER-NOT (SEQ ID NO:     )  
 AAGACTGCAGACCATGGCGG 20-MER
- 25 TER-3 (SEQ ID NO:     )  
 AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC
- ALECUT (SEQ ID NO:     )  
 CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC
- 30 GLYB (SEQ ID NO:     )  
 GAGGGTGTATTCGGTATCGAGTTGCAGGTTTCGTATC
- GLY3 (SEQ ID NO:     )  
 CTCGATACCCATTACACCCTCACGCCTTTCGA

35 *B. Construction of different promoter vectors*

Various promoters were used to optimize expression and to establish constitutive, heat-shock inducibility and senescence enhancement.

*i. Rice actin promoter and 1<sup>st</sup> intron*

Initial vectors (Figures 11 and 12) were constructed from pCOR105 which was subsequently found to contain a 5bp deletion relative to the published sequence which destroys the *AccI* site (GTAGGTAGAC, deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin  
5 sequence in this region (GTAGGTAG) was therefore restored using oligonucleotide NCO-ACT (CTCACCATGGTAAGCTTCTACC TACAAAAAAGCTCCGCA) by replacing the *Bgl*III/*Hind*III fragment with a PCR product, to produce vector **pPQ10.1**.

A rice repetitive element is present in the upstream region of the actin  
10 promoter used in pCOR105; as this may have unpredictable effects on vector expression it was removed from pPQ10.1 by deletion of the *Kpn*I/*Eco*RI fragment (end-filled with T4 polymerase and ligated following digest, restoring *Eco*RI but not *Kpn*I), to produce vector pGT6. The *Hind*III fragment containing the FAE reading frame and nos terminator of pTP3.1 (see Example 2A) was then inserted  
15 into pGT6 to produce construct **pJO6.3**.

*ii. Soybean heat-shock promoter*

A soybean heat shock promoter from a 23kD HSP was obtained from pMA406 (Ainley & Key (1990) PMB 14:949-60). This promoter when fused to  $\beta$ -  
20 glucuronidase (Jefferson et al 1987 EMBO J 6:3901-3907) had previously been shown to be inducible by a 10°C heat-shock and show stable expression for 24-48 hours (data not shown).  $\beta$ -glucuronidase fusions are a sensitive and versatile fusion marker in higher plants. The construction of the co-integration HS vectors is given below.

25

*iii. Senescence enhanced expression (See1) promoter from Lolium multiflorum*

The promoter and signal sequence (including NPIR motif) of the LSee1 gene was amplified from *Lolium multiflorum* cv Tribune with oligonucleotides SEE-NCO  
30 and SEE-VAC, and cloned as an *Asp*718/*Not*I replacement of the promoter region of vector pTP11.1. Following sequencing to screen for PCR artifacts, one of three identical clones was chosen (pUB8.11).

The See1 promoter from maize has been cloned previously and has

EMBL accession number is AX050343. See WO0070061.

The Lolium version of See1 was also cloned previously (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium multiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence inducible when used to drive both  $\beta$ -glucuronidase and the Agrobacterium ipt gene.

An apoplast-targeted derivative was constructed by amplifying the Potato Protease Inhibitor (PPI) motif with primers PPI-AP6 and SEE-ATG, and cloning the product as an NgoMIV/NotI fragment into pUB8.11 (NgoMIV partial digest), to produce vector pJQ5.2. This vector has both the senescence induced promoter and the apoplast target sequence with the gene to be expressed inserted downstream of the apoplast sequence.

#### PCR Primers

**SEE-VAC (SEQ ID NO: \_\_\_)**

15 AACCATGGCGGCCGCGCGCTCGGTGACGGGCCGGAT

**SEE-NCO (SEQ ID NO: \_\_\_)**

TTCGGTACCATGGCCAGGTATAATTATGG

**SEE-ATG (SEQ ID NO: \_\_\_)**

CTGCGCCGGCGAGATGGMCGTGCACAAGGAG

20

#### C. Construction of targeting sequences

In order to examine whether or not the localization of the enzyme would have an effect on the phenolic acid content of the cell wall various signal sequences were utilized. The targeting sequences were added either to the N-terminus or to the C-terminus of the gene of interest.

##### i. N-terminal signal sequences

Six N-terminal signal sequences were utilized:

- (a) The native Aspergillus end of FAE, plus excretion signal [apoplast localisation]

30 This is from the original clone and has the peptide sequence:

MKQFSAKHVLAVVVTAGHALAASTQGI.

- (b) The mature Aspergillus end, with no excretion signal [cytoplasmic localisation]

Peptide sequence is MAAASTQGI (underlined motif is common to all constructs). Truncation of the signal sequence in (a) above was carried out by PCR with mutagenic primer FAE-N5.

5 (c) The barley aleurain signal, including intact NPIR motif [vacuole localisation]

The barley aleurain vacuolar signal sequence (See Figure 13; Swissprot database accession number P05167) was derived entirely from overlapping primers (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 and ALE CAP-3). Following primer annealing at 37°C and extension with T4 DNA polymerase in the presence  
 10 of dNTPs according to manufacturers instructions, PCR with flanking primers ALE-5 and ALE-3 was carried out. The product was 'polished' with T4 DNA polymerase, purified, digested with NotI and cloned into EcoRV/NotI digested pCOR105-nos terminator vector (see above). ALE-3 contains redundancies so  
 15 that clones encoding NPIR or NPGR motifs may be recovered. Two versions of the signal, with and without the vacuole targeting motif, were produced, to give putative vacuolar NPIR and apoplast (NPGR) signal sequences.

#### PCR Primers

**ALE-5 (SEQ ID NO: \_\_\_)**  
 20 GGAATTTCGTAGACAAGCTTACMATGGCCACGCCCGCGTCCT 41-MER

**ALE-3 (SEQ ID NO: \_\_\_)**  
 TATCCATGGCGGCCGCGGGTTCGGTGACGGGCCGGMVCGGGTTGGAGTC  
 GGCGAA 55-MER

**ALE-CUT (SEQ ID NO: \_\_\_)**  
 25 CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer

**ALECAP-5 (SEQ ID NO: \_\_\_)**  
 GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGG  
 ACGCGG 54-MER

**ALECAP-3 (SEQ ID NO: \_\_\_)**  
 30 TCGCCGTCGCCTCCTCCTCCTTCGCCGACT 33-MER

(d) The barley aleurain signal, mutated to a NPGR motif [cytoplasmic localisation]

(e) The rat sialyl transferase golgi targeting motif [golgi localisation]

5 A Golgi targeting vector, **pJQ3.2**, was made by inserting a reading frame encoding the relevant rat sialyl transferase (RST) motif (See Figure 14. RST motif shown to function in plants by Boevink P, Oparika K, Cruz SS, Martin B, Betteridge A, Hawes C, (1998) PLANT JOURNAL 15 441-447 Stacks on tracks: the plant Golgi apparatus traffics on an actin/ER network) into vector pPQ10.1, and replacing the  
10 EcoRI/NotI promoter/signal fragment of **pJO6.3** with the fragment from this vector. Briefly, the RST motif was constructed by annealing oligonucleotides RST-F1A, RST-F1B, RST-F2A and RST-F2B, and amplifying the product with RST-5AD and RST-3A. This product was cloned and sequenced. Clones were found to have a  
15 deletion which was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of products.

PCR primers

**RST-5AD (SEQ ID NO: \_\_\_)**

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

20 **RST-F1A (SEQ ID NO: \_\_\_)**

TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

**RST-F1B (SEQ ID NO: \_\_\_)**

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGGAGAACTTCT  
TTT

25 **RST-F2A (SEQ ID NO: \_\_\_)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCCCTCCAA  
GCCAAGGA

**RST-F2B (SEQ ID NO: \_\_\_)**

CATTTGGAACCTCCTTGGCTTGGAGGGTG

30 **RST-3A (SEQ ID NO: \_\_\_)**

AACCATGGCGGCCGCCATTTGGAACCTCCTTGGCT

**RST-RPT (SEQ ID NO: \_\_\_)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCC  
AAGCCAAGGA

(f) otif [cytoplasmic localisation]

(g) The potato protease inhibitor II (PPI) apoplast motif [apoplast  
localisation]

5

An apoplast targeting reading frame was designed to encode the relevant potato protease inhibitor II (PPI) motif (See Figure 15) and cloned into pJO6.3, to produce vector pJQ4.9. Briefly, the PPI motif was constructed by annealing oligonucleotides PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 and PPI-AP6, and cloning this product as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI promoter/signal fragment of pJO6.3 was then replaced with the equivalent fragment from the modified pPQ10.1 vector.

10

PCR primers

15

PPI-AP1 (SEQ ID NO: \_\_\_)

GGAATTCGTAGACAAGCTTACMATGGMCGTGACACAAGGAGGT

PPI-AP2 (SEQ ID NO: \_\_\_)

GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC

PPI-AP3 (SEQ ID NO: \_\_\_)

20

CCTACCTCCTGATCGTSCCTCGGCCTCCTCTTGCTCGT

PPI-AP4 (SEQ ID NO: \_\_\_)

CCTTGCGCTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG

PPI-AP5 (SEQ ID NO: \_\_\_)

GTGGACGCCAAGGCCTGCACCCCKCGAGTGCGGCAACCTC

25

PPI-AP6 (SEQ ID NO: \_\_\_)

GGAATTCGCGGCCGCGGGCAGATGCCGAAGCCGAGGTTGCCGCACT

*ii. C-terminal end signal sequences*

Four C-terminal signal sequences were utilized:

30

(a) Native *Aspergillus* end, [CTW] (vacuole and apoplast vectors)

This was derived directly from the genomic clone (see Example 1) as a NcoI-SphI fragment (Sph end filled with T4 polymerase) which replaces the



Nco1-Not1 region of a standard actin -FAE vector (Not1 end filled with T4 DNA polymerase).

(b) Expression vector linker alone [CTW-PVAAA] (plant optimised C-terminus for vacuole, golgi and apoplast vectors)

5 CTW is the peptide sequence of the Aspergillus FAE COOH end and is here provided by oligo FAE3. In this primer the reading frame is extended to provide the additional amino acids PVAAA which are partially encoded by the Not1 site used for cloning downstream signals see c) and d) below. Some COOH amino acids /motifs may affect compartment targeting, the PVAAA  
10 sequences are expected to be neutral in this respect while the native Aspergillus end may not be.

(c) Linker plus KPLKDEL [first K is primer artifact, intended to be E] (ER retention vectors)

15 These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone. KDEL targeting has been demonstrated in plants by Denecke et al. ((1992) EMBO J 11: 2345-2355 Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope).

(d) Linker plus ETTEG [frameshift of (c)] (loss of ER retention - vacuole  
20 vectors)

These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone (see Example 2A).

25 The KDEL signal is for ER retention, while others provide controls. A frameshift in the TER5 region [additional A] was used in subsequent constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA.

#### *D. Co-integration and co-transformation vectors.*

##### *30 Co-transformation vectors*

A Hygromycin resistance gene driven by a CaMV345S promoter (**pRob5**) (35S-HYG-CMV in pUC18 (modified HYG, derived from pGL2) Bilang et al (1991)

Gene 100:247-50) was used for co-transformation experiments with **pTT3** and **pTP3.1, pJQ4.9, pJQ3.2, pJO6.3, pJQ5.2, pUB8.1** 1 vectors.

5 *Co-integration vectors*

1. **Actin promoter constructs - pTR2.22, pTR6.1, pTR8.1, pTR9.4, pTR7.1, pTT5.5 and 5.1.**

10 The CAMV35S-hyg region from pAJEB64TCA [a plant expression vector constructed by Andy Bettany at IGER containing CaMV-HYG from pTRA151 (Zheng et al 1991 Plant Physiol 97:832-835) (CaMV35S-HYG-tnI terminator as clonable cassette in pUC4) cloned into KpnI site of pCOR105] was added as a HindIII fragment at the KpnI site (T4 polymerase blunt) of pTP4a2, in divergent orientation  
15 to FAE to create **pTR2.22**. The FAE/Nos HindIII fragment of this vector was replaced as follows in co-expression vectors. From pTP5.1 for **pTR6.1**, from pTP10.1 to **pTR8.1**, from pTP11.1 to **pTR9.4**. Signal sequences of FAE in **pTR2.22**  
20 were replaced as HindIII/BglII fragments in **pTR7.1** (fragment from pT09.1). PCR products (ALE5/ALE-G) was digested with AccI and T4 polymerase, polished, followed by NotI digest and cloning into EcoRV/ NotI digested pTR2.22 to give clones **pTT5.5** and **5.1**.

PCR primer

**ALE-G**

TATCCATGGCGGCCGCGCGGTTCGGTGACGGGCCGGCCCGGGTTGGAGTC  
25 GGCGAA

2. **Actin promoter constructs -pUF1, pUA1K3, pUH4, pUH5, pUH6, pUH7, pUH8, pUH9.**

30 The HygR gene from pAJEB64TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give **pHOX3**. For ease of cloning the downstream HindIII site was destroyed to create **pUA1K3** and replacement of the FAE/Nos terminator HindIII fragment in this vector  
35 was carried out as follows. From pTP5.1 for **pUF1**, from pTP11.1 for **pUH4**, from pTP8.5 for **UH5**, from pTT5 for **pUH6**, from pUA4.4 for **pUH7**, from pTU5 for **pUH8** and from pUG4 for **pUH9**.

### 3. Heat-shock promoter constructs - pUH10, pUH12, pUC5.11.

5 A co-transformation vector in which FAE is expressed from the soybean heat shock promoter was made by first modifying pMA406 to remove the nos terminator (BglII linearised and gel purified, KpnI digested, T4 DNA polymerase polished in the presence of dNTPs and recircularised), and then inserting the FAE HindIII fragment from pTP11.1, creating pTT3.1, which encodes the full aleurain signal and the native  
10 *Aspergillus* COOH-terminus.

Following assays of various constructs, co-integration vectors were constructed with FAE and HygR genes arranged in tandem.

The HygR gene from pAJEB-64-TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1,  
15 to give pHOX3 and subsequently excised as a HindIII/SacI fragment (partial SacI digest, relevant sites found in flanking pTP3.1 sequences) which was cloned into the HindIII/SacI sites of pMA406, in tandem orientation (vector pUH1a20). FAE sequences were then cloned into the HindIII site of pUH1a20 downstream of the heat-shock promoter (HindIII fragment from pTU5 for pUH10, HindIII fragment  
20 from pTT5 for pUH12). A pTP3.1 derivative was made by cloning the CaMV/HygR HindIII cassette from pAJEB-64-TCA in tandem orientation downstream of the FAE gene in pTP3.1, inactivating the middle HindIII site by partial digestion and end-filling, and excising the combined FAE/HygR cassette as a single HindIII fragment, which was inserted at the HindIII site in pMA406 to  
25 produce pUC5.11.

#### Example 3

##### Transformation of Plant Cells

Eight to ten weeks old embryogenic *F. arundinacea* and *L. multiflorum*

30 suspension cultures were bombarded either with a single co-integration plasmid DNA vector containing FAE and hyg resistance genes, or with a co-transformation vector containing FAE and with plasmid pROB5 conferring hygromycin resistance (CAMV35S-hpt- nos) using a Particle Inflow Gun (PIG) (Finer et al. (1992) Development of the particle inflow gun for DNA delivery to

plant cells Plant Cell Reports 11:323-328) and 1.5-3.0  $\mu\text{m}$  gold particles as in Dalton *et al* (Dalton et al. (1999) Co-transformed diploid *Lolium perenne* (Perennial ryegrass), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Damel) plants produced by microprojectile bombardment. Plant Cell Reports. 5 18: 721-726) and Kuai et al (Regeneration of fertile transgenic tall fescue (*Festuca arundinacea*) plants with a stable highly expressed foreign gene. Plant Cell Tissue and Organ Culture (1999) 58:149-154). Transformants were selected with hygromycin (25 to 50mg/l) over a 10-12 week selection period at 25°C under continuous white fluorescent light ( $60 \mu\text{E m}^2 \text{s}^{-1}$ ) and plants regenerated via 10 somatic embryogenesis as in Dalton *et al* 1999, *supra*. Regenerated plants were screened for FAE activity on transfer to soil and expressing plants grown to maturity in a containment growth room at 18°C under 16h fluorescent lights ( $350 \mu\text{E m}^2 \text{s}^{-1}$ ). Mature plants (6-8 weeks old) were re-assayed for FAE activity and fresh tissue harvested for Southern, Northern and Western analysis, and for self 15 digestion analysis. The remaining tissue was freeze dried and powdered for cell wall structure analysis, In vitro-dry matter digestibility (IVDMD) determinations and for in-vitro gas production determinations of rates of tissue digestion.

#### Example 4

##### Targeting of Expression Product

20 To verify that the targeting sequences are effective in delivering the gene the targeting sequences were operably linked to a green fluorescent protein GFP. The vector constructs are shown in Figure 16. Cells were transformed by particle bombardment as in Example 3. Localization of the GFP could be 25 visualized under a microscope 1 day after bombardment (i.e., shooting). See Figure 16.

#### Example 5

##### FAE1 activity

30 Plants regenerated from transformed cells showed FAE activity in all plant tissues tested. Cells were transformed as above under the direction of the ER and APO targeting sequences. FAE activity in transformed *Festuca arundinacea*

leaves of different ages was elevated compared to control (untransformed) plants. See Figures 17 and 18.

Similar results were seen with *Lolium mutiflorum* leaves at different ages transformed as above under the direction of vacuolar, ER and APO targeting sequence. See Figures 19 and 20.

FAE expression under a heat shock promoter can also be induced. (Data not shown.)

Thus, we have demonstrated FAE expression in *Festuca* and *Lolium* leaves under constitutive and HS promoters with effective FAE targeting to the vac, ER and apo.

#### *FAE assay*

FAE activity was determined in soluble extracts of fresh (or frozen at -70°C) leaves or cell cultures (0.5g) with 0.1M NaAc, pH 5.0 extraction buffer. Extracts were incubated with 24mM EF (ethyl 4-hydroxy-3-methoxycinnamate) or 1% FAXX as substrate, at 28°C for 24hrs and FAE activity calculated as the amount of ferulic acid released. FAE activity was also determined by measuring the release of monomeric and dimeric ferulic acid from self-digested leaf or cell culture samples. Fresh, or frozen, leaves or cell cultures (0.5g) were ground in 0.1M NaAc, pH5.0 extraction buffer in the presence and absence of xylanase (1000U GC140/sample) without added substrate and incubated at 28°C for 72hrs. Following incubation, and centrifugation, soluble extracts were loaded onto an activated reverse phase C18  $\mu$ Nova sep-pak column (Waters), eluted with 100% MeOH and the MeOH sample analysed by HPLC.

25

#### **Example 6**

##### **Chemical Analysis of Cell Wall Extracts**

Ester bound compounds were extracted from freeze dried powdered leaves or cell cultures (50 -100mg) with NaOH (5ml of 1M) followed by incubation at 25°C for 23hrs under N<sub>2</sub>. After centrifugation and acidification of the soluble extract with concentrated HCl, the extracted phenolics were loaded onto an activated reverse phase C18  $\mu$ Nova sep-pak column (Waters) and eluted with 100% MeOH. and the MeOH sample analysed by HPLC.

30

HPLC was carried out with methanol: 5% acetic acid either with a 35-65% MeOH gradient in 15min (FAE assay) or with a 30-70% MeOH gradient in 25 min (monomer and dimer cell wall components) at 2ml/min on a  $\mu$ Nova Pak C18 8x10 RCM (Waters). Extracts were detected and quantified with a diode array detector  
5 (240-400nm Waters 996PDA) monitored at 280nm for aldehydes and 340nm for hydroxycinnamic acids. .

Levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under VAC, and ER and APO targeting sequences are reduced compared to control (untransformed) plants.  
10 The results can be seen in Figure 21 and 22, respectively. Thus, we show where this does not result in reduced cell wall phenolics in growing plants with vac targeting but does result in lower phenolics with ER and apo targeting. In addition,

Levels of esterified monomeric and dimeric hydroxycinnamic acids in  
15 *Festuca arundinacea* plants expressing FAE are not significantly reduced when FAE is VAC, targeting (Fig 21) which is as predicted for correct vacuolar targeting, but are significantly reduced, as predicted, in some plants when FAE was ER and APO targeted, compared to control (untransformed) plants. The results can be seen in Figure 22.

20

#### Example 7

##### *In vitro* dry matter digestibility. (IVDMD)

The *in vitro* dry matter digestibility (IVDMD) was estimated on 1.0 g dry weight of leaf or cell culture tissue using the pepsin/cellulase method of Jones  
25 and Hayward (The effect of pepsin treatment of herbage on the prediction of dry matter digestibility from solubility in fungal cellulase solutions. Journal of the Science of Food and Agriculture (1975) 26:711-718).

We show that the presence of FAE in the plants results in higher digestibility of the leaves. This may be due to internal FAE activity acting on  
30 normal cell walls with vacuole located FAE and to both FAE activity and the lower cell wall crosslinking with ER and apo targeted FAE (as also found with cell cultures).

End point digestibility as determined by IVDMD were higher in leaf tissue of some transformed plants of *Festuca* expressing FAE, compared to control (untransformed) plants. Examples are shown where vacuolar, ER or apoplast targeted FAE under a constitutive actin promoter have been effective at increasing IVDMD. Similar results were obtained with in leaves of *Lolium*, but were less pronounced.

The results can be seen in Figures 23 and 24.

### Example 8

#### *In vitro* gas production measurements

In each experiment, 1.0-g samples of freeze dried powdered leaf tissue or cell culture were fermented in three 165-ml capacity serum bottles according to the pressure transducer technique of Theodorou et al. (Theodorou et al. (1994) A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Animal Feed Science and Technology* 48: 185-197). Grab samples of rumen-digesta were taken at 8.00 h before the morning feeding from fistulated wethers fed grass hay, and transported to the laboratory in a pre-warmed (39°C) vacuum flask. The microbial inoculum and culture media were prepared as described by Theodorou et al. (1994). Each serum bottle received 10 ml of microbial inoculum, 85 ml of buffer and 4 ml of reducing agent.

At the end of the incubation period, (144h) the contents of each serum bottle were filtered through pre-weighed sintered glass funnels and freeze dried to constant weight. Dry matter loss was calculated as the difference between the dry weight of the sample pre- and post-incubation. Additionally, the concentration of volatile fatty acids (VFA) in the liquid fraction of the culture media at the end of the 144-h incubation period was determined by gas chromatography. A Chrompack CP 9000 chromatograph fitted with an automatic sampler (Chrompack 911) and a flame-ionisation detector, linked to a Dell PC with A1-450 integration software, was used for VFA quantification.

Gas production data were fitted to the model of France et al. (France, J., Dhanoa, M.S., Theodorou, M.K, Lister, S.J., Davies. D.R. and Isac, D. 1993. A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*. 163: 99-111.) using the MLP (Ross,

G.J.S. 1987. *MLP, Maximum Likelihood Program Version 3.08*. Oxford Numerical Algorithms Group) package. The equation is in the form,  $Y = A\{1 - e^{[-b(t-T) - c\sqrt{t-T}]}\}$ , where  $Y$  is the cumulative gas production (ml),  $A$  is the asymptote (i.e. gas pool),  $T$  is lag time, and  $b$  ( $h^{-1}$ ) and  $c$  ( $h^{-0.5}$ ) are decay rate constants. A combined fractional rate ( $h^{-1}$ ) of gas production ( $\mu$ ) was calculated as,  $\mu = b + c/2\sqrt{t}$ , where  $t$  is the incubation time (h).

It can be seen for *Festuca arundinacea* (denoted as BN in Figure 25) that cell cultures have a higher rate of digestion and cumulative gas production in the presence of FAE and that the addition of an exogenous xylanase further enhance the availability of fermentable carbohydrates. Similar results are found in FAE expressing cultures without added FAE. Fermentation rates are further increased compared with controls by the addition of exogenous FAE or xylanase as these cultures expressing FAE have a reduced cell wall phenolic composition to controls Figures 26-28.

#### Example 9

##### FAE & xylanase transformed plants

Addition of exogenous xylanase (GC140) greatly increased FAE mediated release of phenolics from *Festuca* and *Lolium* leaves expressing *A. niger* FAE. See Figures 29-31 which show that phenolic release from leaf cell walls is increased in all FAE expressing plants on cell death and this is stimulated by xylanase irrespective of the targeting. Therefore expression of a fungal xylanase in plant cells is tested.

The FAE expression cassette is modified to comprise a fungal xylanase gene (either *T. reesei* or *A. niger*) to yield a FAE-xylanase expression cassette. The FAE-xylanase expression cassette is used to transform plant cells in a manner similar to those described in Example 3. The transformed cells are allowed to grow and are selected on an appropriate medium. The enzymes so expressed increase the availability of fermentable carbohydrates to a greater extent than the FAE expression cassette.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will



be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

What is claimed:

1. A transgenic plant comprising an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.  
5
2. The plant of claim 1, wherein the polynucleotide is derived from *Aspergillus niger*.
3. The plant of claim 2, wherein the polynucleotide is FAE I from *Aspergillus niger*.
- 10 4. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with an altered glycosylation site.
5. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with a substitution so that glycosylation is altered.
6. The plant of claim 3, wherein the polynucleotide further  
15 comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
7. The plant of claim 3 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
8. The plant of claim 1, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by sexual reproduction.
- 20 9. The plant of claim 1, wherein the promoter is an inducible promoter.
10. The plant of claim 9, wherein the promoter is a senescence promoter.
11. The plant of claim 9, wherein the promoter is a heat shock  
25 promoter.
12. The plant of claim 1, wherein the promoter is a constitutive promoter
13. The plant of claim 1, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the

polynucleotide.

14. The plant of claim 13, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

5 15. The plant of claim 14, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene

16. The plant of claim 15, wherein the targeted gene is a barley aleurain gene.

10 17. The plant of claim 15, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum or apoplast signal sequence.

18. The plant of claim 15, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted senescence gene.

19. The plant of claim 18, wherein the senescence gene is a *Lolium See1* signal sequence.

15 20. The plant of claim 13, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.

21. The plant of claim 20, wherein the targeted gene is a rat sialyl transferease signal sequence.

20 22. The plant of claim 13, wherein the polynucleotide is derived from the signal sequence of an apoplast signal sequence.

23. The plant of claim 22, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

24. The plant of claim 16, wherein the polynucleotide is derived from *Solanum tuberosum*.

25 25. The plant of claim 13, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

26. The plant of claim 25, wherein the polynucleotide sequence is a KDEL sequence.

27. The plant of claim 25, wherein the polynucleotide sequence is a stop codon.
28. The plant of claim 25, wherein the polynucleotide sequence is an extension of the ferulic acid esterase reading frame to provide a linker to  
5 KDEL.
29. The plant of claim 1, further comprising introduction into the plant a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.
30. The plant of claim 29, wherein the xylanase encoding  
10 polynucleotide is from *Trichoderma reesei*.
31. The plant of claim 29, wherein the first and second expression cassettes are present on separate plasmids.
32. The transgenic plant of claim 1, selected from the group consisting of Festuca, Lolium, Zea and Avena.
- 15 33. The transgenic plant of claim 32, wherein the plant is a Festuca plant.
34. A method of controlling the level of phenolic acids in plant cell walls of a transgenic plant, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a ferulic acid  
20 esterase encoding polynucleotide.
35. The method of claim 34, wherein the polynucleotide is derived from *Aspergillus niger*.
36. The method of claim 35, wherein the polynucleotide is a FAE  
1 gene from *Aspergillus niger*.
- 25 37. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with an altered glycosylation site.
38. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with a substitution such that glycosylation is altered.
39. The method of claim 36, wherein the polynucleotide

comprises CTWPVAAA at the 3' end.

40. The method of claim 36 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.

41. The method of claim 36, wherein the polynucleotide  
5 comprises SEQ ID NO:1.

42. The method of claim 34, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by transformation of cell cultures.

43. The method of claim 42, wherein the cell cultures are  
10 regenerated to plants.

44. The method of claim 34 wherein the ferulic acid esterase polynucleotide is introduced into the plant by sexual reproduction.

45. The method of claim 34, wherein the transgenic plant is a member of a genus selected from the group consisting of *Festuca*, *Lolium*,  
15 *Avena* and *Zea*.

46. The method of claim 45, wherein the transgenic plant is a member of the genus *Festuca*.

47. The method of claim 46, wherein the transgenic plant is a *Festuca arundinacea*.

48. The method of claim 34, wherein the promoter is an  
20 inducible promoter.

49. The method of claim 48, wherein the promoter is a senescence promoter.

50. The method of claim 48, wherein the promoter is a heat  
25 shock protein promoter.

51. The method of claim 34, wherein the promoter is a constitutive promoter.

52. The method of claim 51, wherein the promoter is an actin

promoter.

53. The method of claim 34, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the polynucleotide.

5 54. The method of claim 53, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

55. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene.

10 56. The method of claim 55, wherein the targeted gene is a barley aleurain gene.

57. The method of claim 55, wherein the polynucleotide is derived from the signal sequence of a *Lolium See1* signal sequence.

15 58. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum signal sequence.

59. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce an apoplast signal sequence.

20 60. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.

61. The method of claim 60, wherein the targeted gene is a rat sialyl transferease signal sequence.

25 62. The method of claim 59, wherein the polynucleotide is derived from the signal sequence of a fungal apoplast signal sequence.

63. The method of claim 62, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

64. The method of claim 59, wherein the apoplast signal

sequence is from a potato.

65. The method of claim 53, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

5 66. The method of claim 65, wherein the polynucleotide sequence is a KDEL sequence.

67. The method of claim 65, wherein the polynucleotide sequence is a stop codon.

68. The method of claim 65, wherein the polynucleotide  
10 sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.

69. The method of claim 34, further comprising simultaneous introduction into the plant a second expression cassette comprising a promoter operably linked to a polynucleotide encoding a xylanase gene.

15 70. The method of claim 69, wherein the second polynucleotide is a fungal xylanase.

71. The method of claim 70, wherein the fungal xylanase is from *Trichoderma reesei*.

72. The method of claim 35, wherein the first and second  
20 expression cassettes are present on separate plasmids.

73. The method of claim 1, wherein the first and second expression cassettes are present on separate plasmids.

74. A transgenic plant produced by the method of claim 34.

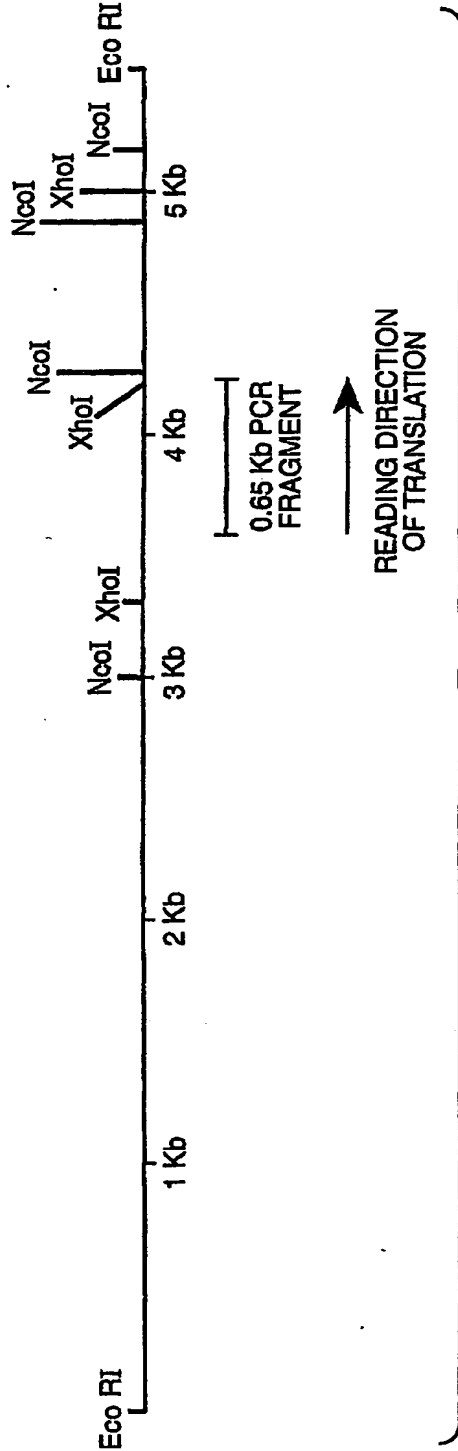


FIG. 1



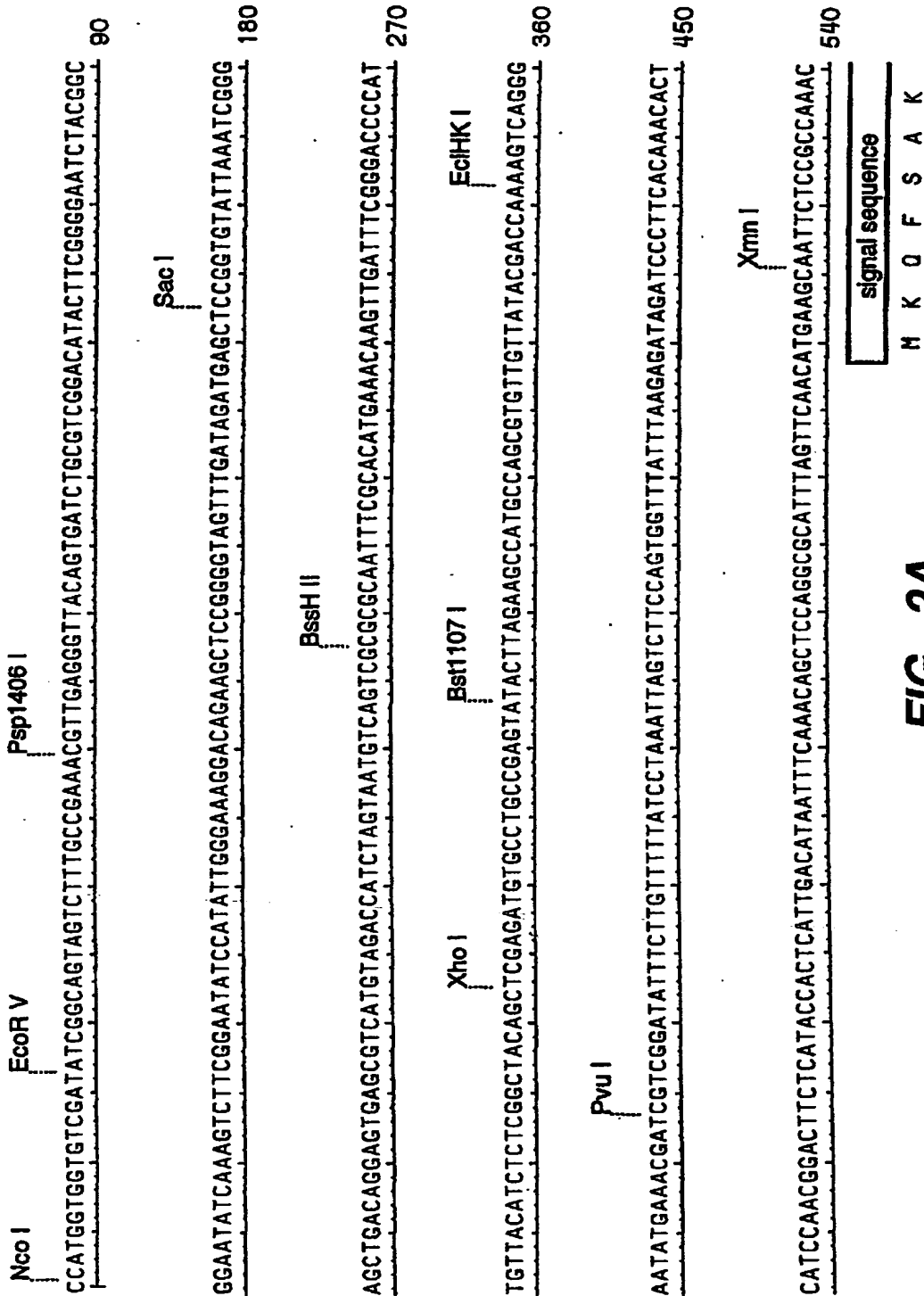


FIG.-2A

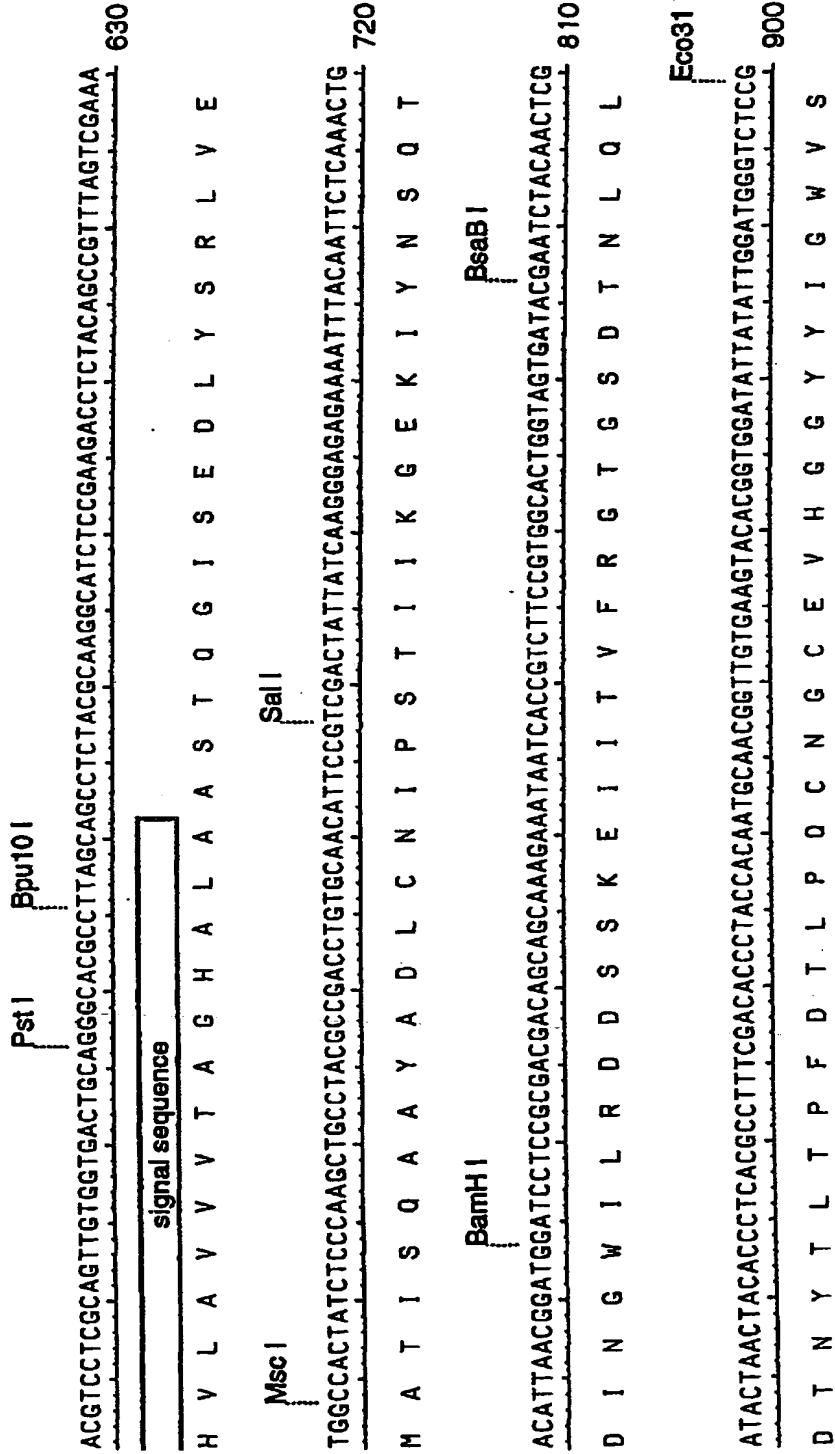


FIG.-2B

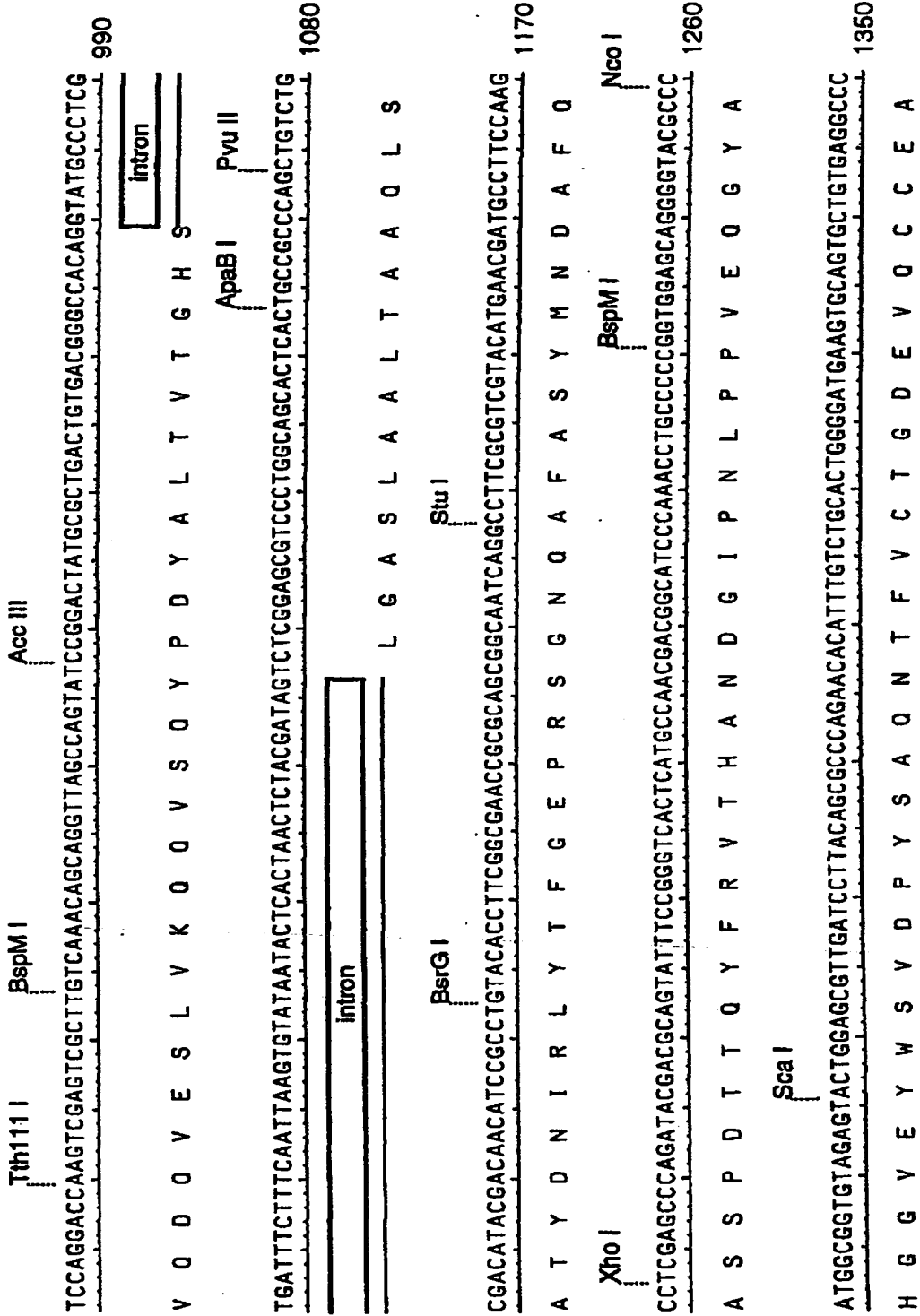


FIG.-2C

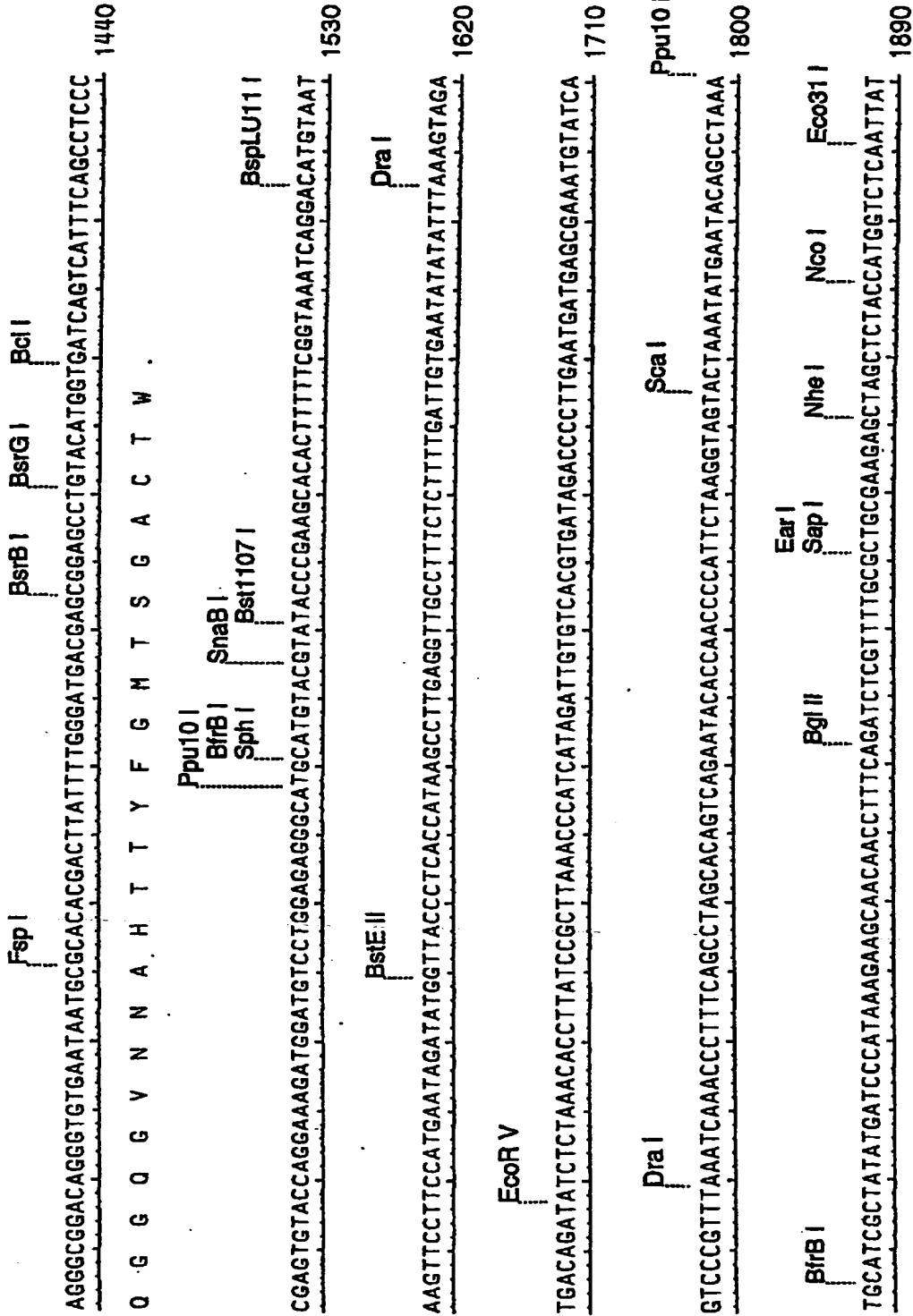


FIG.-2D

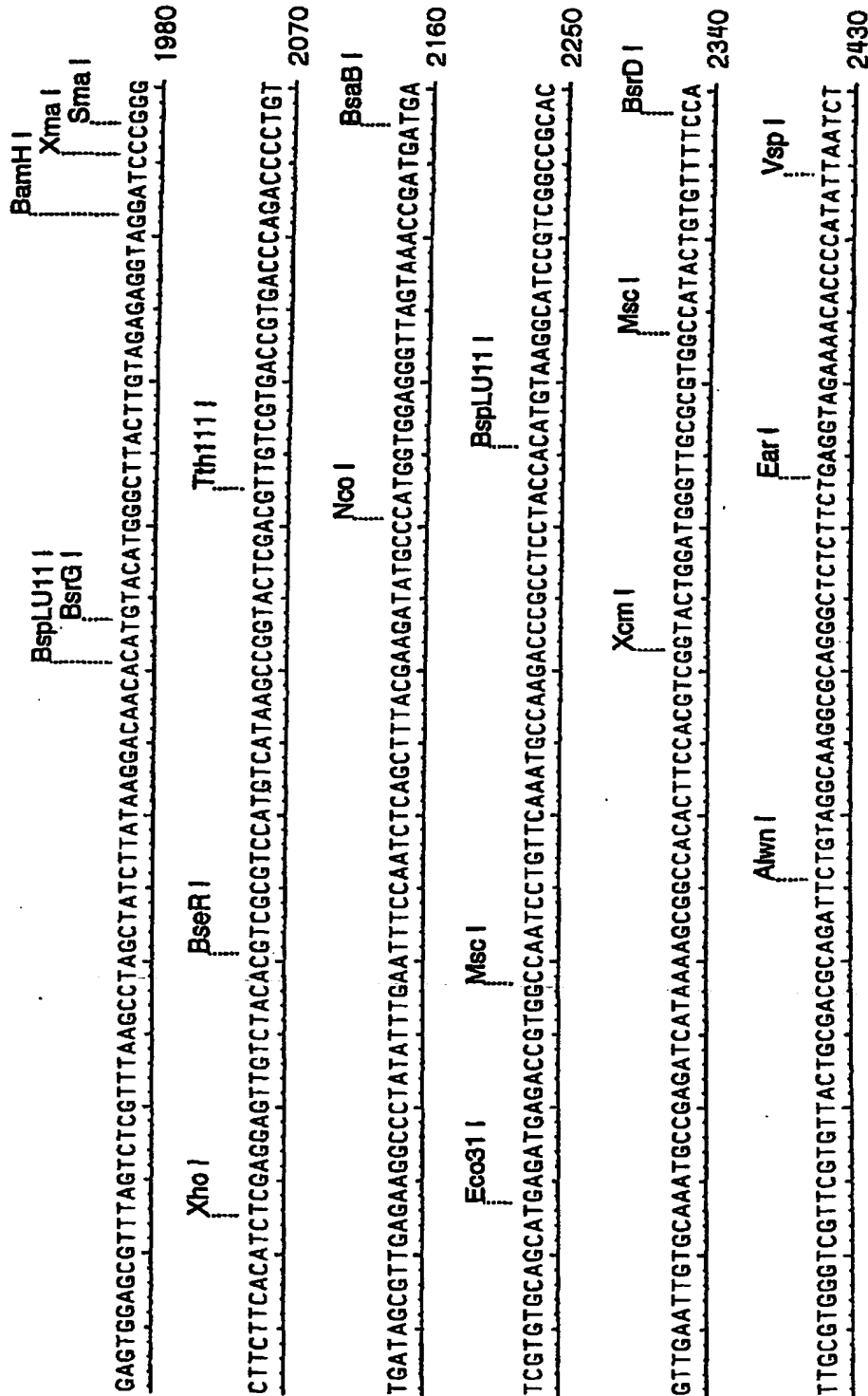


FIG.-2E

CCATGGTGTGTCGATATCGGCAGTAGTCTTTGCCGAAACGTTGAGGGTTACAGTGATCTGCCGTCGGACATACTT  
 CGGGGAATCTACGGCGGAATATCAAAAGTCTTCGGAAATATCCATATTTGGAAAGGACAGAAAGCTCCGGGGTAGTTT  
 GATAGATGAGCTCCGGTGTATTAATCGGGAGCTGACAGGAGTGAGCGTCATGTAGACCATCTAGTAATGTCAGT  
 CGCGCGCAATTTCCGACATGAAACAAGTTGATTTCCGGACCCCATTTGTACATCTCTCGGCTACAGCTCGAGATG  
 TGCCCTGCCGAGTACTCTTAGAAGCCATGCCAGCGTGTGTTATACGACCACAAAGTCAAGGAAATATGAAAACGATCG  
 TCGGATATTTCTTGTTTTATCCTAAATAGTCTTCAGTGGTTTATTTAAAGAGATAGATCCCTTCACAAAACACT  
 CATCCAAACGGACTTCTCATACCACTCATTTGACATAAATTTCAAACACAGCTCCAGGGCGCATTTAGTTCAACATGAAGC  
 AATTCCTCCGCCAAACACGCTCTCGCAGTGTGTTGTTGACTGCGAGGACCGCTTAGCAGCCCTCTACGCAAGGCATCT  
 CCGAAGACCTCTACAGCCGTTTAGTTCGAAAATGGCCACTATCTCCAAAGCTGCTACGCCGACCTGTGCAACATTC  
 CGTCGACTATTAACAAGGAGAGAAAATTTACAATTTCTCAAACCTGACATTAACGGATGGAATCCTCCGGACGACA  
 GCAGCAAAGAAATAATCACCCGCTTCCGGTGGCACTGGTAGTACGAAATCTACAACTCGATCTAACTACACCC  
 TCAGCCCTTTCGACACCCCTACCAAAATGCAACGGTGTGAAAGTACACGGTGGATATTATATTGGATGGGTCTCCG  
 TCCAGGACCAAGTCGAGTCCGTTGTCAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC  
 ACAGGTATGCCCTCGTGATTTCTTTCAATTAAGTGTATAAATACTCACTAACTCTACGATAGTCTCGGAGCGTCCC  
 TGGCAGCACTCACTGCCGCCAGCTGTCTGCGACATACGACAACATCCGCCCTGTACACCTTCGCGGAAACCCGCCA  
 GCGGCAATCAGGCCCTTCGCGCTGTACATGAACGATGCCCTTCCAAAGCTCGAGCCACAGATACGACGCGAGTATTTCC  
 GGGTCACTCATGCCAACGACGGCATCCCAAACCTCCCGGTGGAGCAGGGGTACGCCCATGCGGGTGTAGGT  
 ACTGGAGCGTTGATCCCTTACAGCGCCCAAGAACACATTTGTCTGCACTGGGATGAAGTGCAGTCTGTGAGGCC  
 AGGGGGACAGGGTGTGAAATAATGCGCACACGACTTATTTTGGGATGACGAGCGGAGCCGTGTACATGGTGTACAG  
 TCATTTCAAGCCCTCCCGAGTGTACCAGGAAAGATGGATGCTCCGGAGAGGCA TGCA TGTACGTATACCCGAAAGC  
 ACACCTTTTCGGTAAATCAGGACATGTAATAAGTTCCTTCCATGAAATAGATATGGTTACCCCTCACCATAAAGCCTT  
 GAGGTTGCCCTTCTTCTTTGATTTGATTAATAATAATAATAAGTAGATGACAGATACTCTAAACACCTTATCCGCT  
 TAAACCCATCATAGATTGTGTCAAGTGTGAAATAATAATAATAAGTAGATGACAGATACTCTAAACACCTTATCCGCT  
 AACCTTTCAGCTAGCACAGTCAAGATAACCAACCCCATTTCAAGGTAGTACTAAATAATAATAACAGCCCTAAA  
 TGCA TCGCTAATATGATCCCATAAAGAGCAACAACTTTTCAGATCTCGTTTGGCTGGAGGAGCTAGCTCTAC  
 CATGGTCTCAATATGAGTGGAGCGTTTAGTCTCGTTTAAAGCTTAGCTATCTTAATAAGGACAAACACATGTACATG  
 GGCTTACTTGTAGAGAGGTAGGATCCCGGCTTCTTCAATCTCGAGGAGTTGTTCTACACGTCGCGCTCCATGTCA  
 TAAAGCCGGTACTCGACGTTGTCTGACCGTGACCCAGACCCCTGTTGATAGCGTTGAGAAAGGCCCTATATTTGAA  
 TTTCCAAATCTCAGCTTACGAAGATATGCCCATGTTGGAGGGTTAGTAAACCGATGATGATCGTGTGCGAGCATGA  
 GATGACCGGTGGCCAACTCTGTTCAAAATGCCAAGACCCGCTTCCATCCACATGTAAGGCACTCCGTCGGCCGAC  
 GTTGAATGTGCAAAATGCCGAGATCATAAAAGCGGCCACACTTCCACGTCGGTACTGGATGGGTTGCCGCTGGCC  
 ATACTGTGTTTCCATGCGGTGGTCTGTTGTTACTGCGACCGCAGATTTCTGTAGGCAAGGCCAGGGCTCTCT  
 TCTGAGGTAGAAAACACCCCATATAATACTGAAATTC

FIG.-3

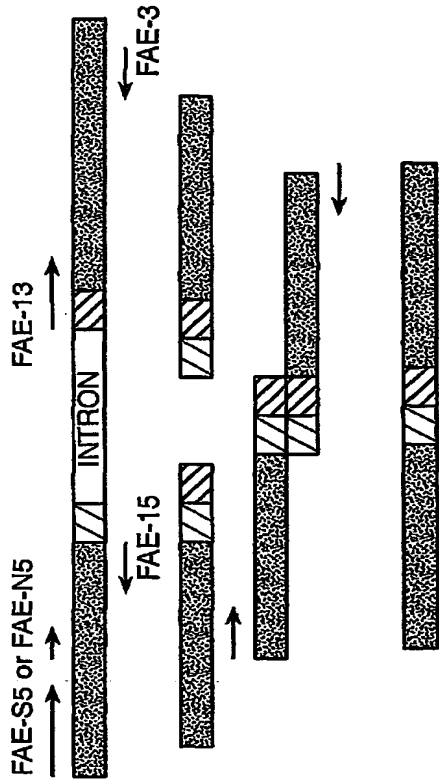


FIG. 4

FAE-13 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCACTC 35-mer  
 FAE-15 GCGCCGAGGAGTGGCCGGTCACGGTCAAGCGGTAGTCC 40-mer

intron position in original

i

Y A L T V T G H S L G A S L A A L  
 G G A C T A C G G C T G A C C G T G A C C G G C C A C T C C C T C G G G C C  
 C C G G C C A C G C C C T C G G C C T C C C T G G C G G C A C T C  
 Y A L T V T G H A L G A S L A A L  
 complement, FAE-15  
 FAE-13

FIG. 5

Vector Construction

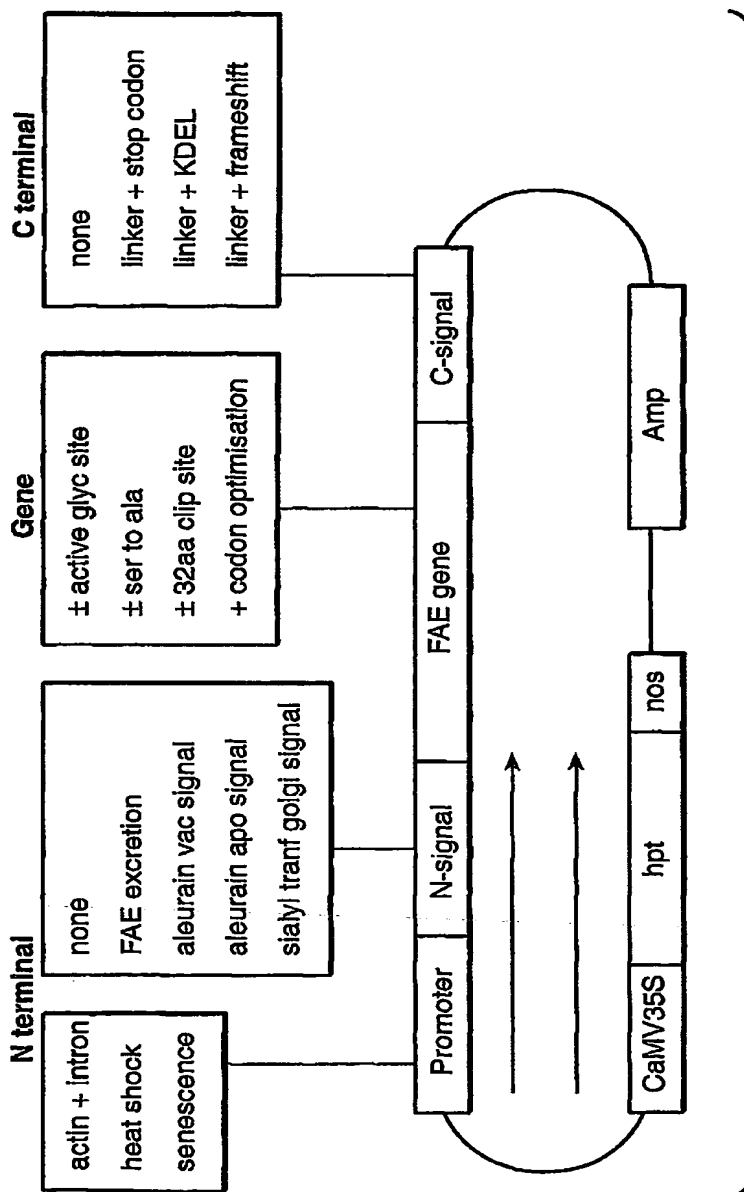
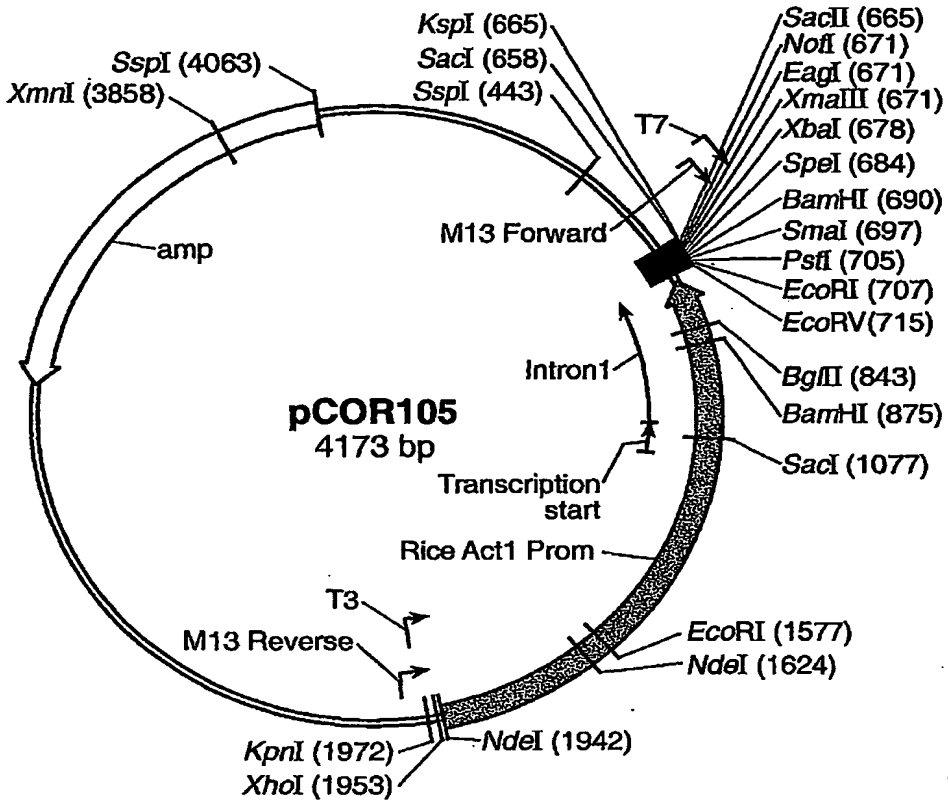
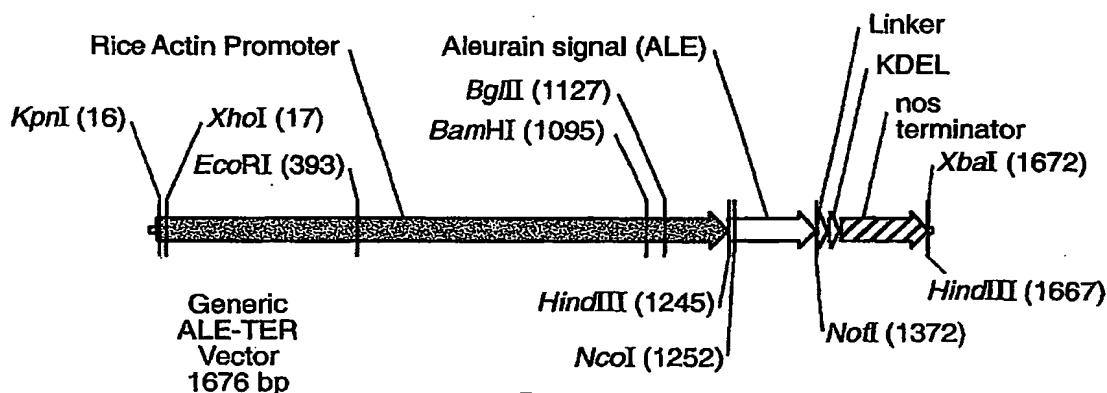


FIG.-6





**FIG. 7**



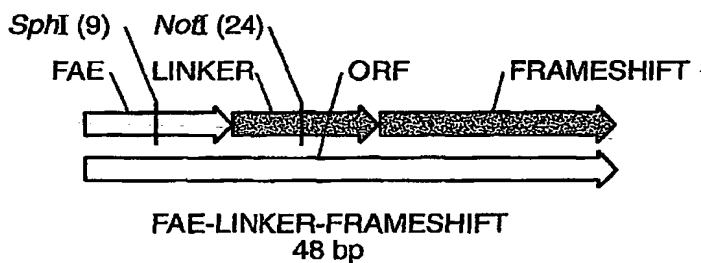
**FIG.\_8**

KDEL-COOH ER retention sequence

NotI  
 -----  
 A A A K P L K D E L \*  
 1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

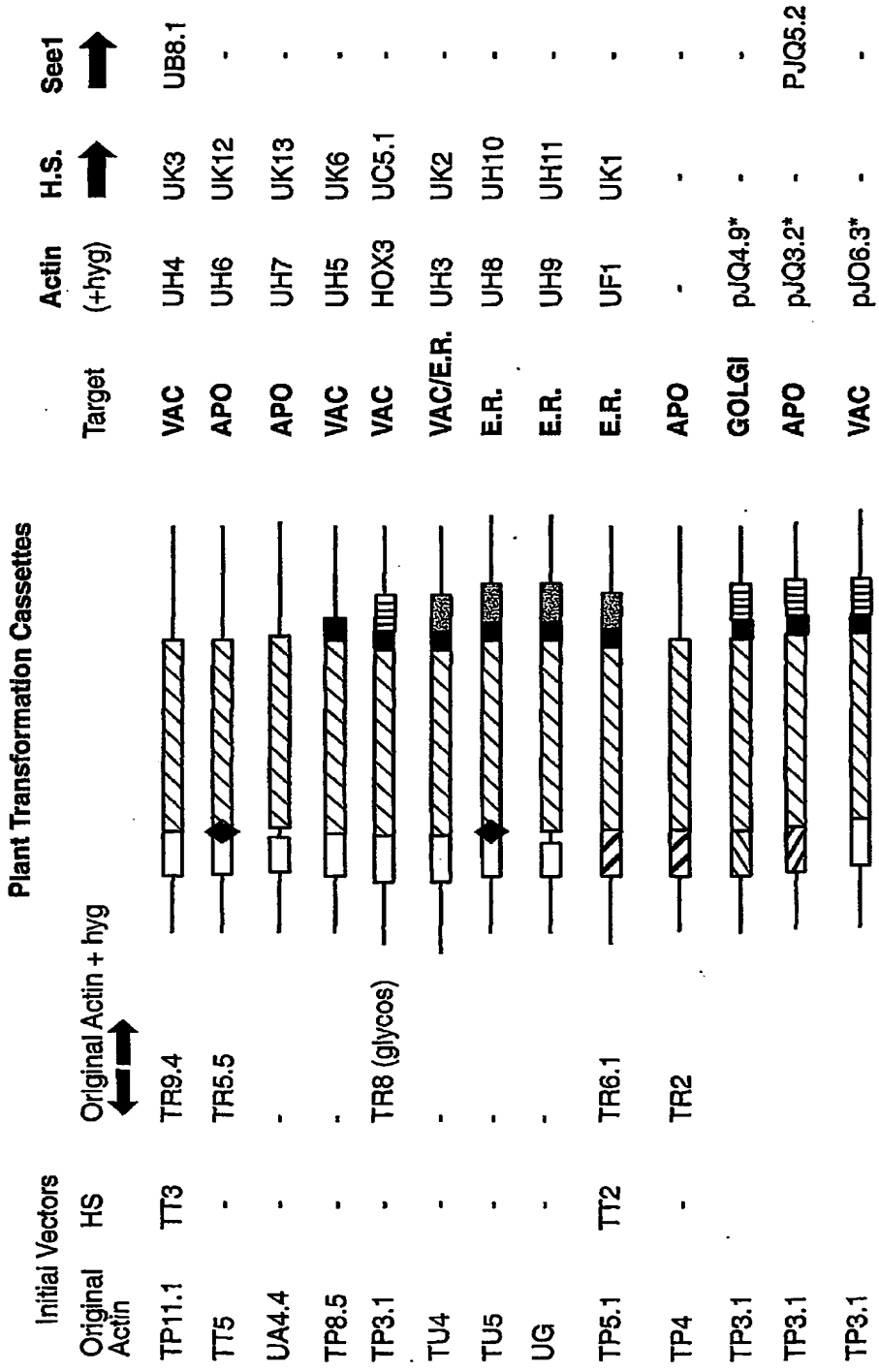
**FIG.\_9**

F AE-LINKER-FRAMESHIFT Structure and Sequence

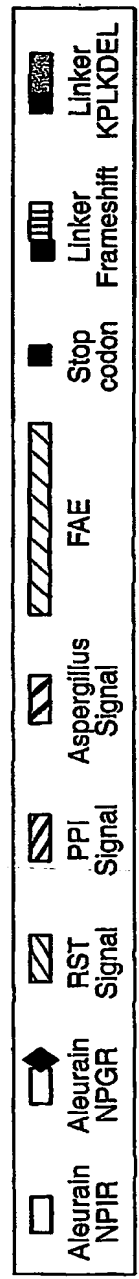


+1 G A C T W P V A A A E T T E G  
 SphI NotI  
 -----  
 1 GCGCATGCA CCTGGCCGGT CGCGGCCGCG GAAACCACTG AAGGATGA  
 CCGGTACGT GGACCGCCA GCGCCGGCGC CTTGGTGAC TTCCTACT

**FIG.\_10**



\* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)



**FIG.-11**

**Vectors****Original Actin promoter in pCOR105**

	<b>Target</b>	<b>Signal sequences</b>	<b>Vectors</b>
(i)	APO	- aleurain-NPGR-FAE - aleurain-delNPIR -FAE	pUH6, pTT5, TT5.5, pTT5.1 pUH7, pUA4.4,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL - aleurain-delNPIR-FAE-linker-KDEL	pTU5, pUH8, pUG4, pUH9,
(iii)	VAC	- aleurain-NPIR-FAE	pTP11.1, pTR9.4, pUH4, pUK3,
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pTU4, pUH3,
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUA1K3, pTP3.1, pUC5.11
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pTP8.5, pUH5
(vii)	ER	- Aspergillus signal -FAE-KDEL	pTP5.1, pTP6.1, pUF1,

**Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)**

(i)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pJ06.3
(ii)	GOLGI	- RST-FAE-linker-frameshift	pJQ3.2
(iii)	APO	- PPI-FAE-linker-frameshift	pJQ4.9

**Heat-shock promoter**

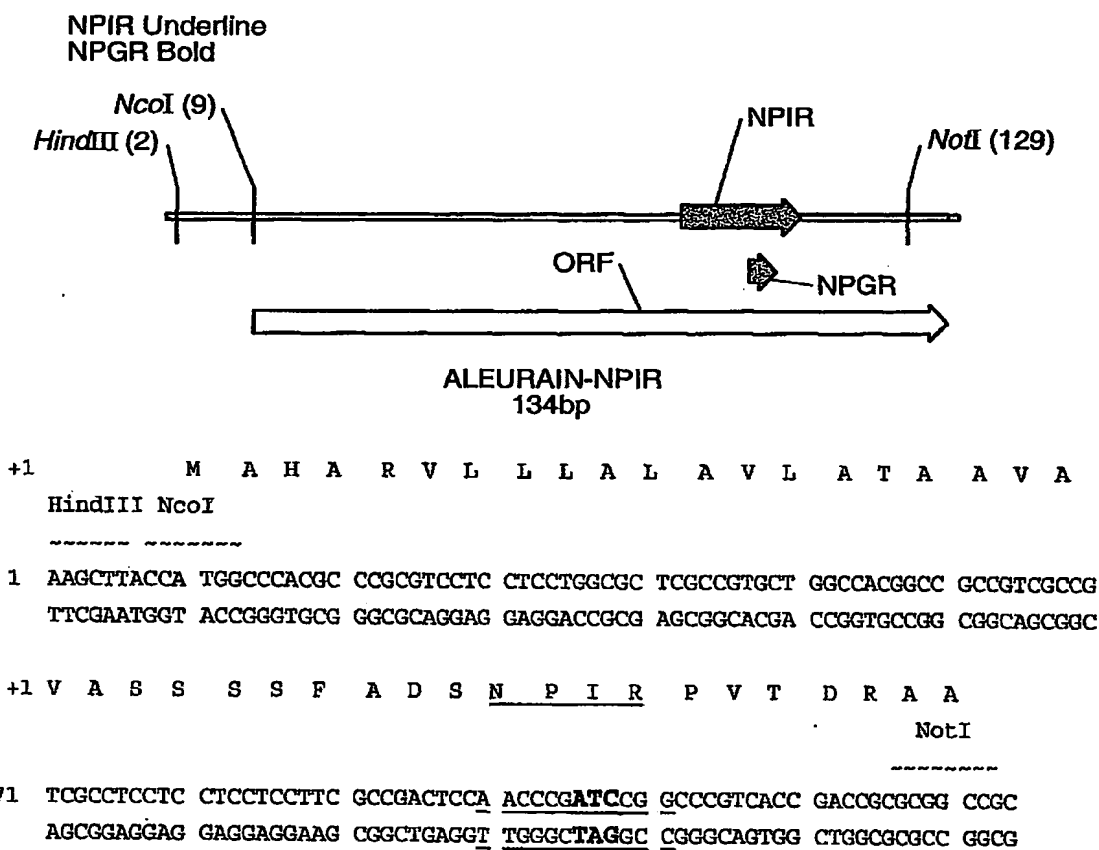
(i)	APO	- aleurain-NPGR-FAE - aleurain-delNPIR-FAE - Aspergillus signal-FAE	pUH12 pUH13 pTP4a2, pTR2.22,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL - aleurain-delNPIR-FAE-linker-KDEL	pUH10 pUH11
(iii)	VAC	- aleurain-NPIR -FAE	pUK3, pTT3
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pUK2
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUC5.11, pHOX3
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pUK6
(vii)	ER	- Aspergillus signal -FAE-KDEL	pUK1, pTT2

**Senescence promoter**

(i)	APO	- See1-PPI-FAE-linker-frameshift	pJQ5.2
(ii)	VAC	- See1-aleurain-deleted NPIR-FAE	pUB8.1

**FIG. 12**

**ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast)  
Structure and Sequence**



**FIG. 13**

RAT SIALYL TRANSFERASE Golgi signal sequence

HindIII  
 ~~~~~  
 M I H T N L K K K F S L F I L V F L L F A  
 1 AAGCTTACCA TGATCCACAC CAACCTCAA AAGAAGTTCT CCCTCTTCAT CCTCGTCTTC CTCTCTTTCG  
 . V I C V W K K G S D Y E A L T L Q A K E F Q M  
 71 CCGTGATCTG CGTGTGAGAG AAGGCTCCG ACTACGAGGC CCTCACCTTC CAAGCCAAGG AGTTCCAAT

NotI  
 ~~~~~  
 . A A  
 141 GCGGCGCGC

FIG.-14

POTATO PROTEASE INHIBITOR II Apoplast signal sequence

HindIII  
 ~~~~~  
 M X V H K E V N F V A Y L L I V L G L L L  
 1 AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSTC  
 GGCCTCCTCT

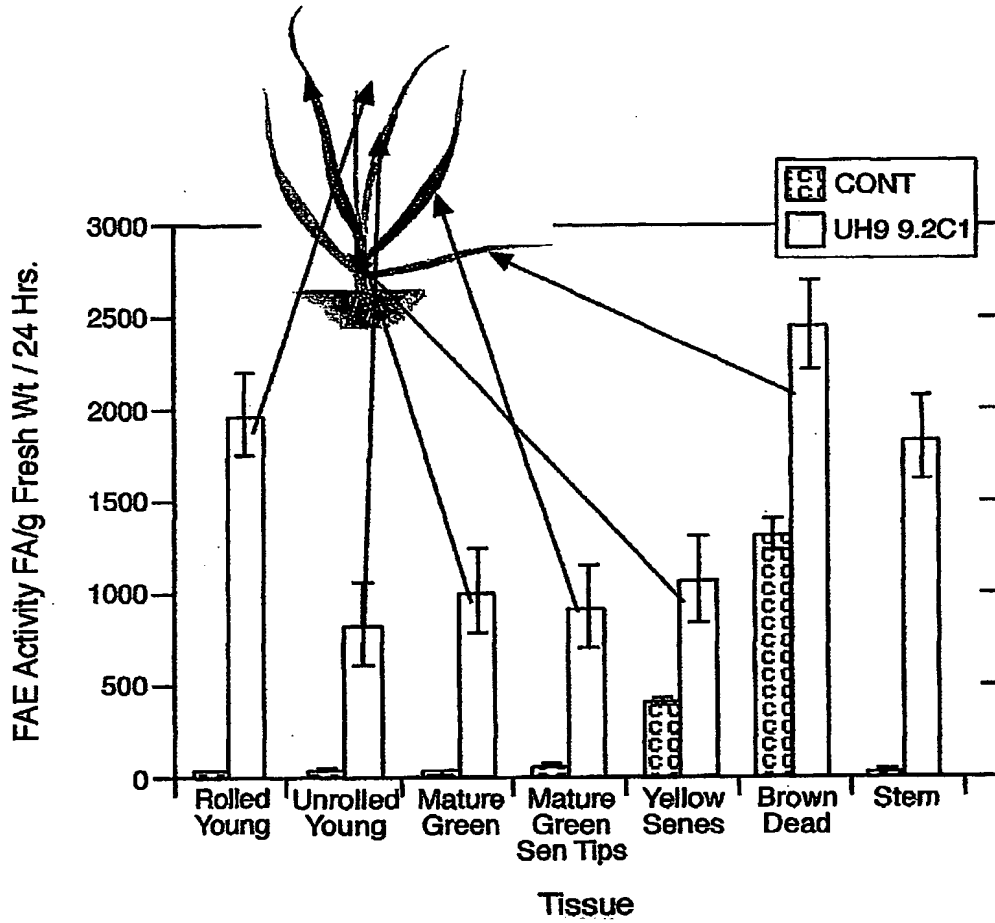
NcoI  
 ~~~~~  
 . L V S A M E H V D A K A C T X E C G N L  
 G F G .  
 71 TGCTCGTSTC CGCCATGGAG CACGTGGAGC CCAAGGCCCTG CACCKCGAG TGGGGCAACC  
 TCGGCTTCGG

NotI  
 ~~~~~  
 . I C P A A A  
 141 CATCTGCCGG GCGGCGGCC

FIG.-15

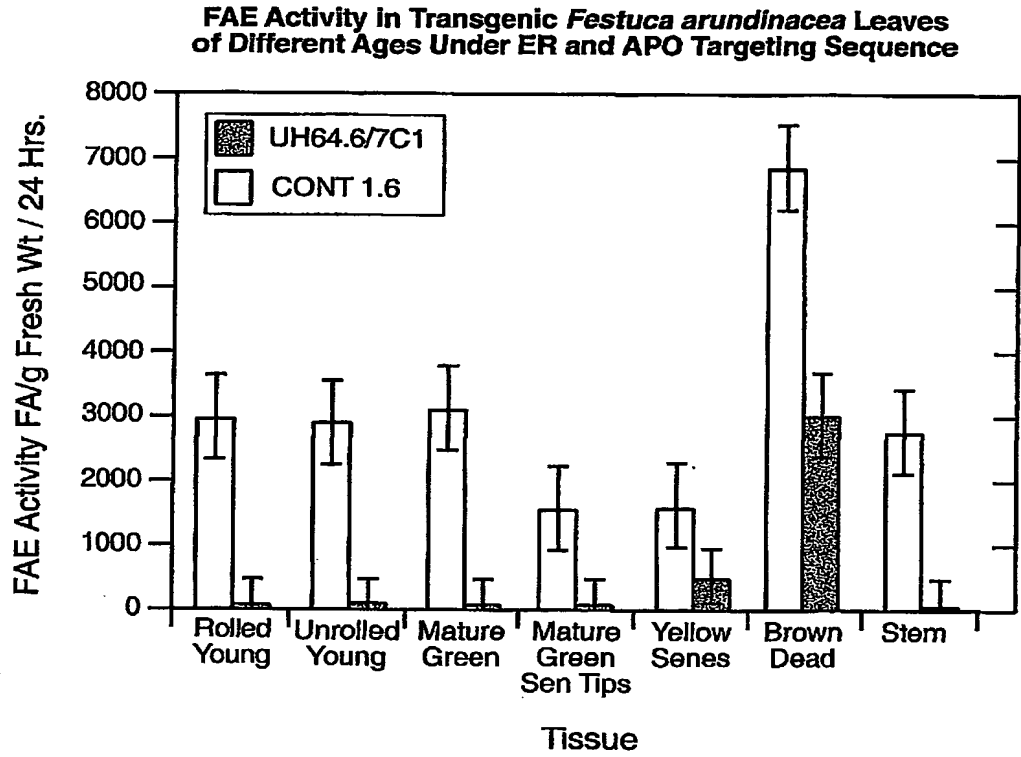


**FAE Activity in Transgenic *Festuca arundinacea* Leaves of Different Ages Under ER and APO Targeting Sequence**

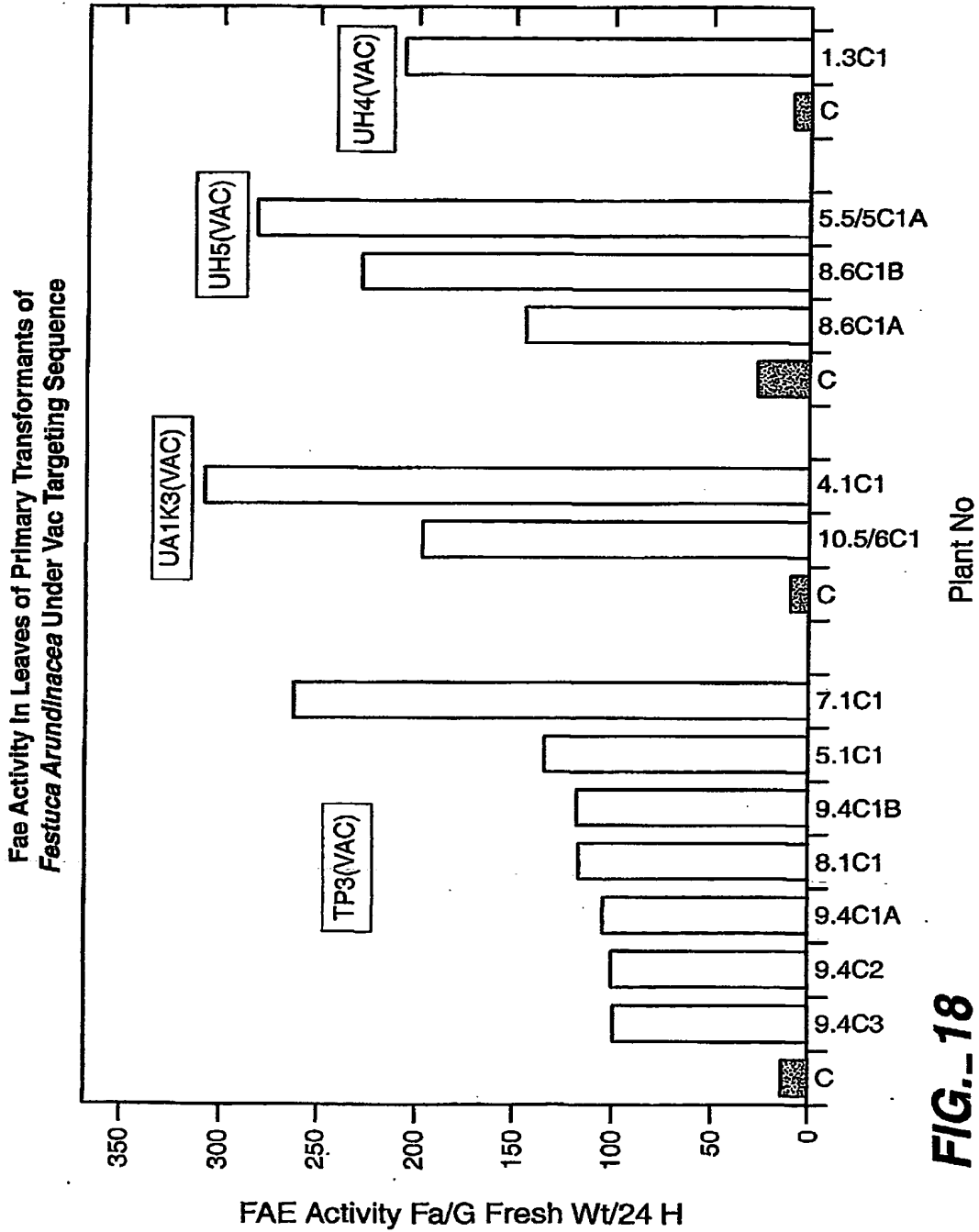


**FIG. 17A**



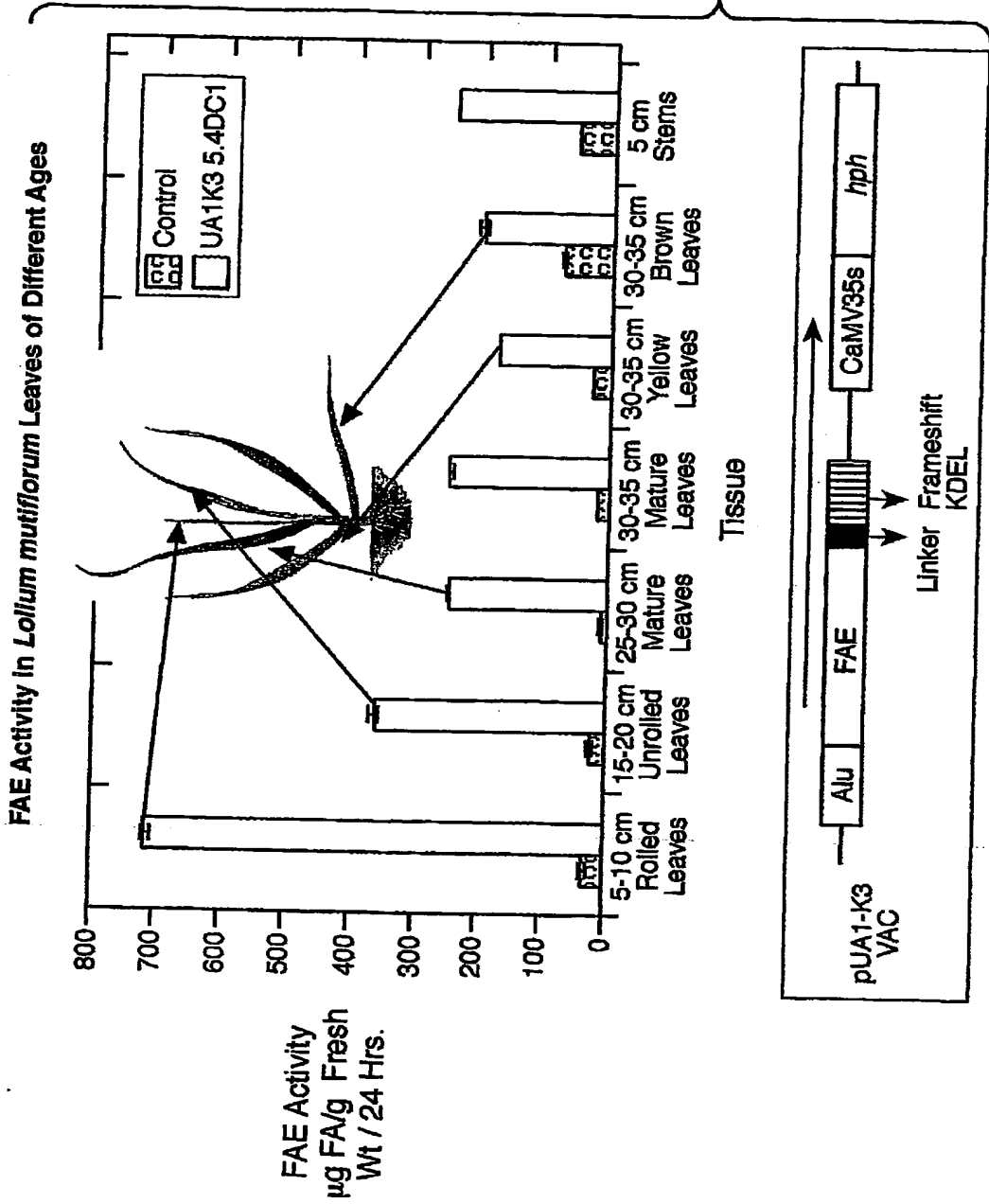


**FIG. 17B**



**FIG. 18**

FIG. 19



FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

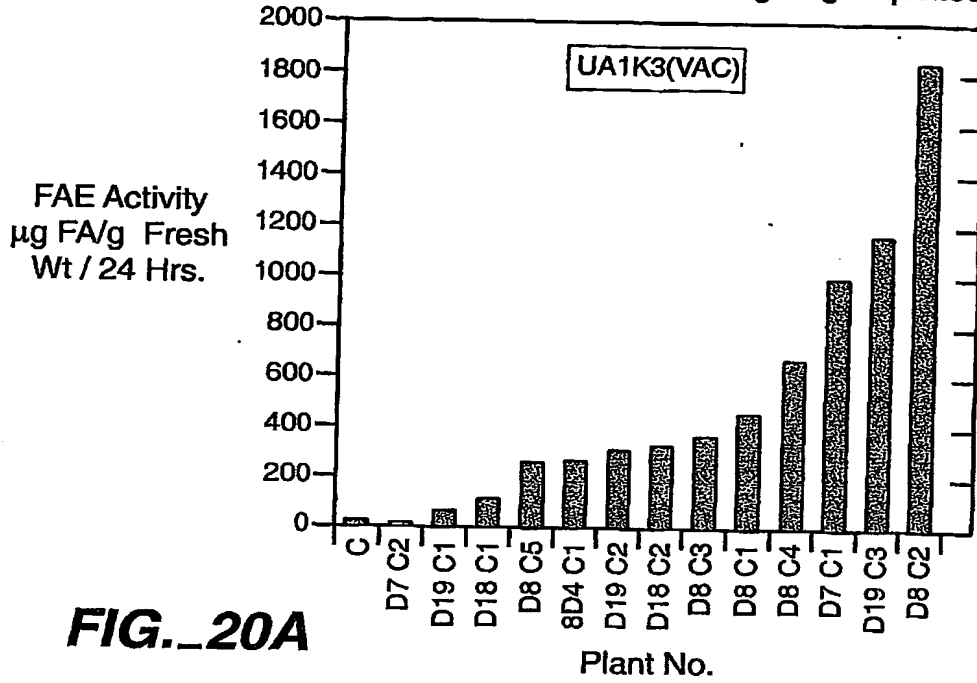


FIG. 20A

FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

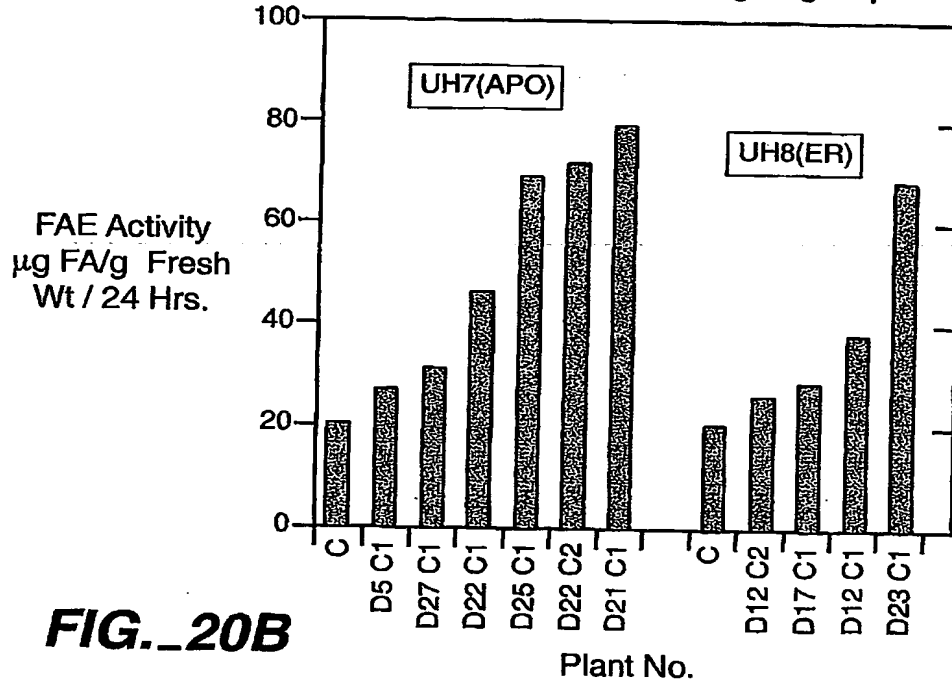


FIG. 20B

Levels of Esterified Monomeric and Dimeric Hydroxycinnamicacids In *Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

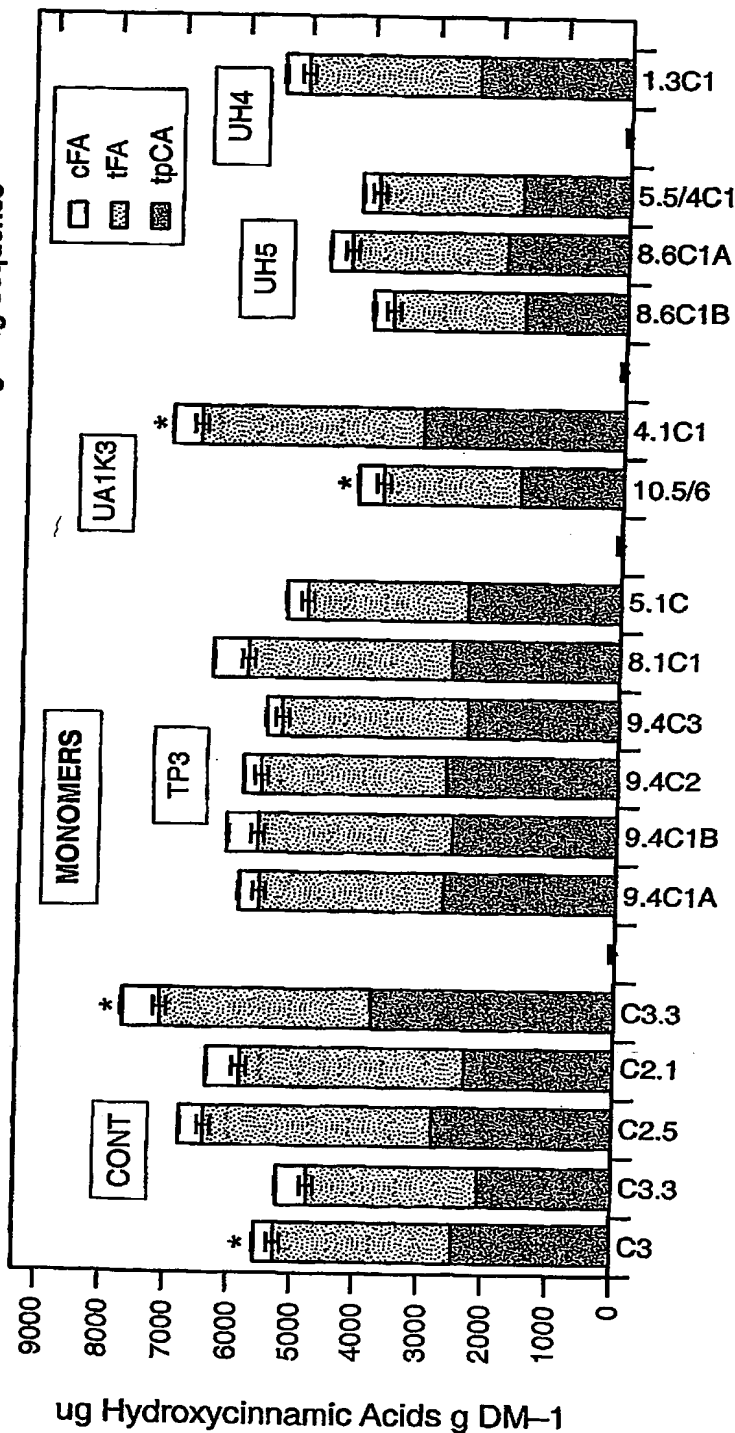
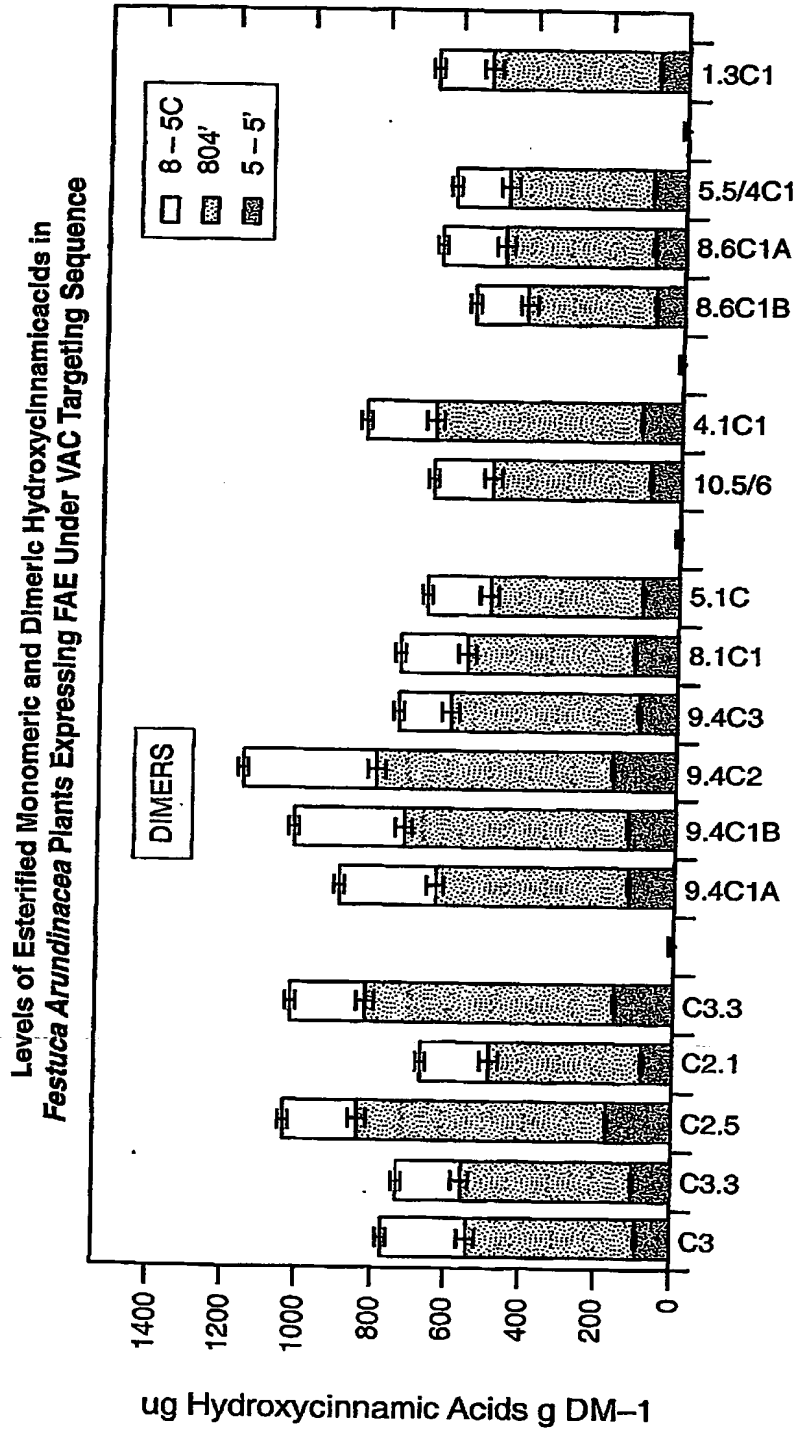


FIG. 21A



**FIG. 21B**

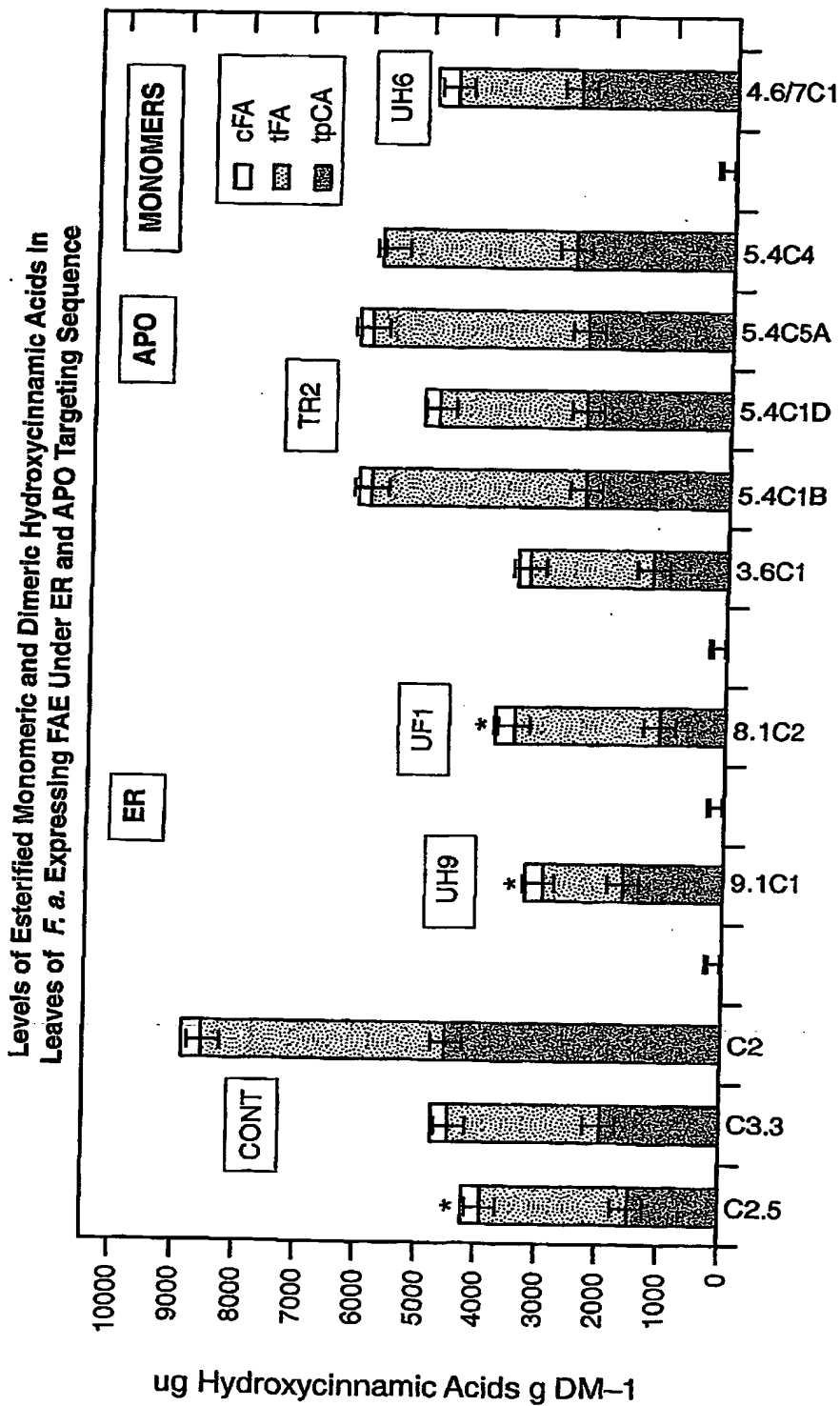
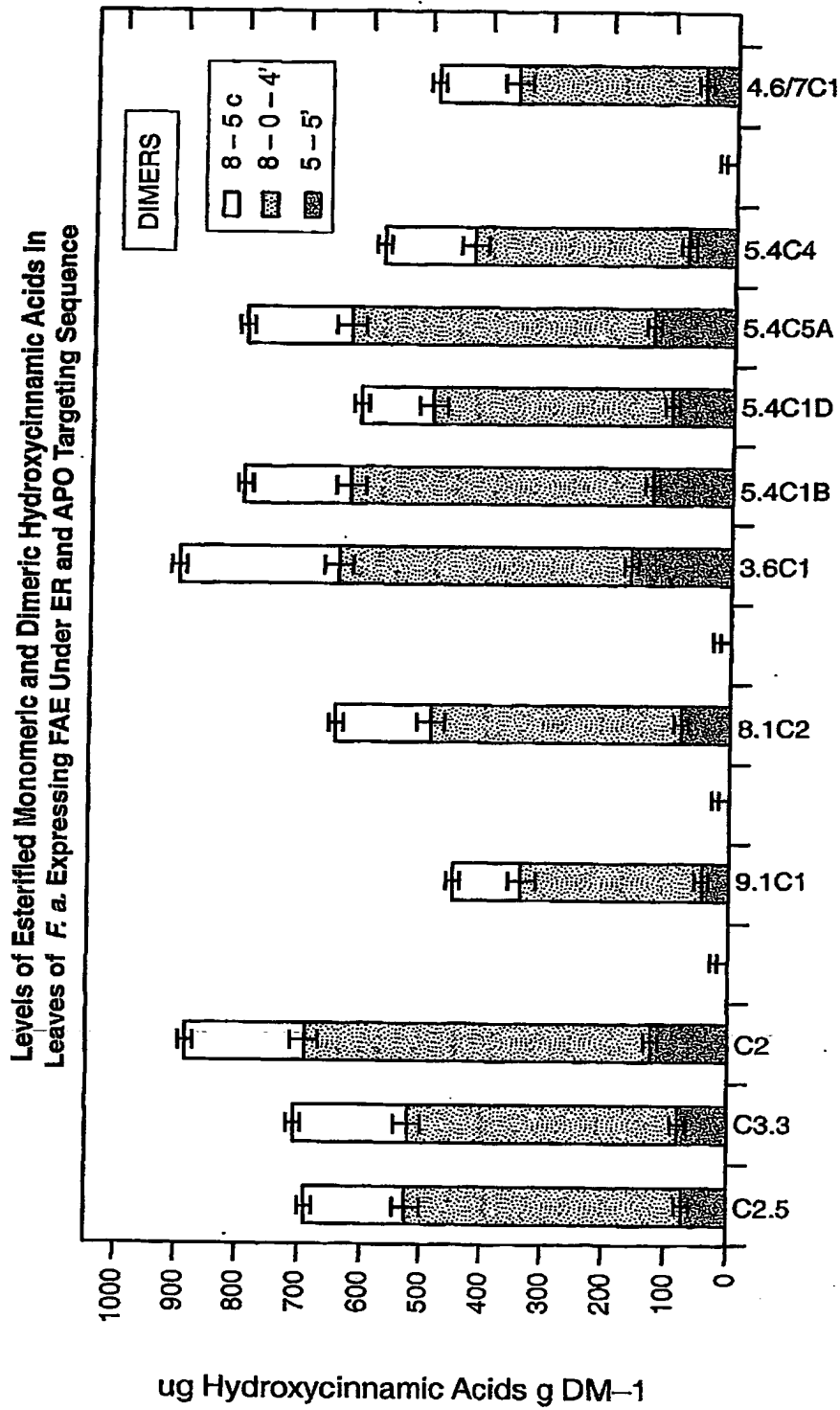
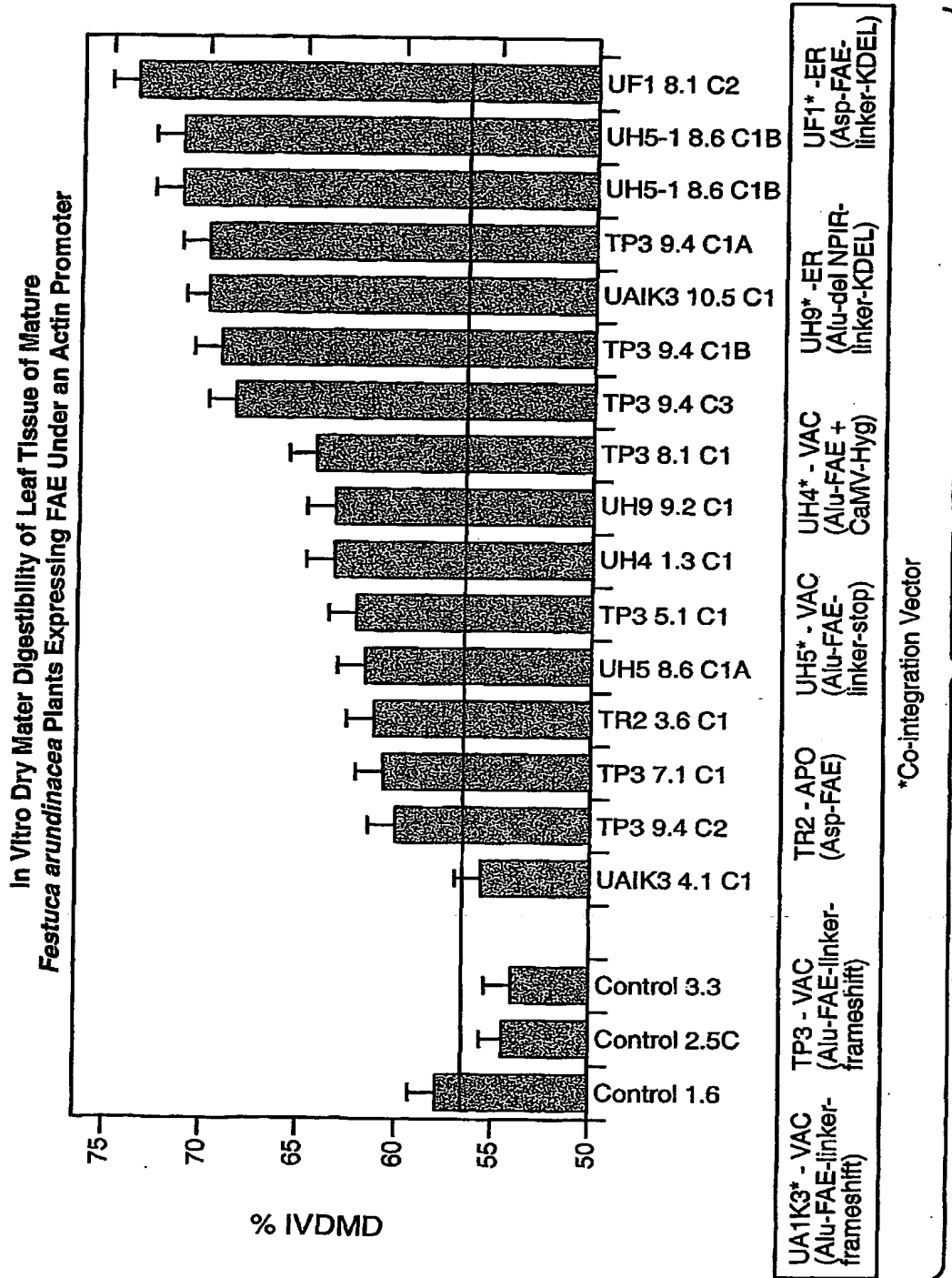


FIG. 22A

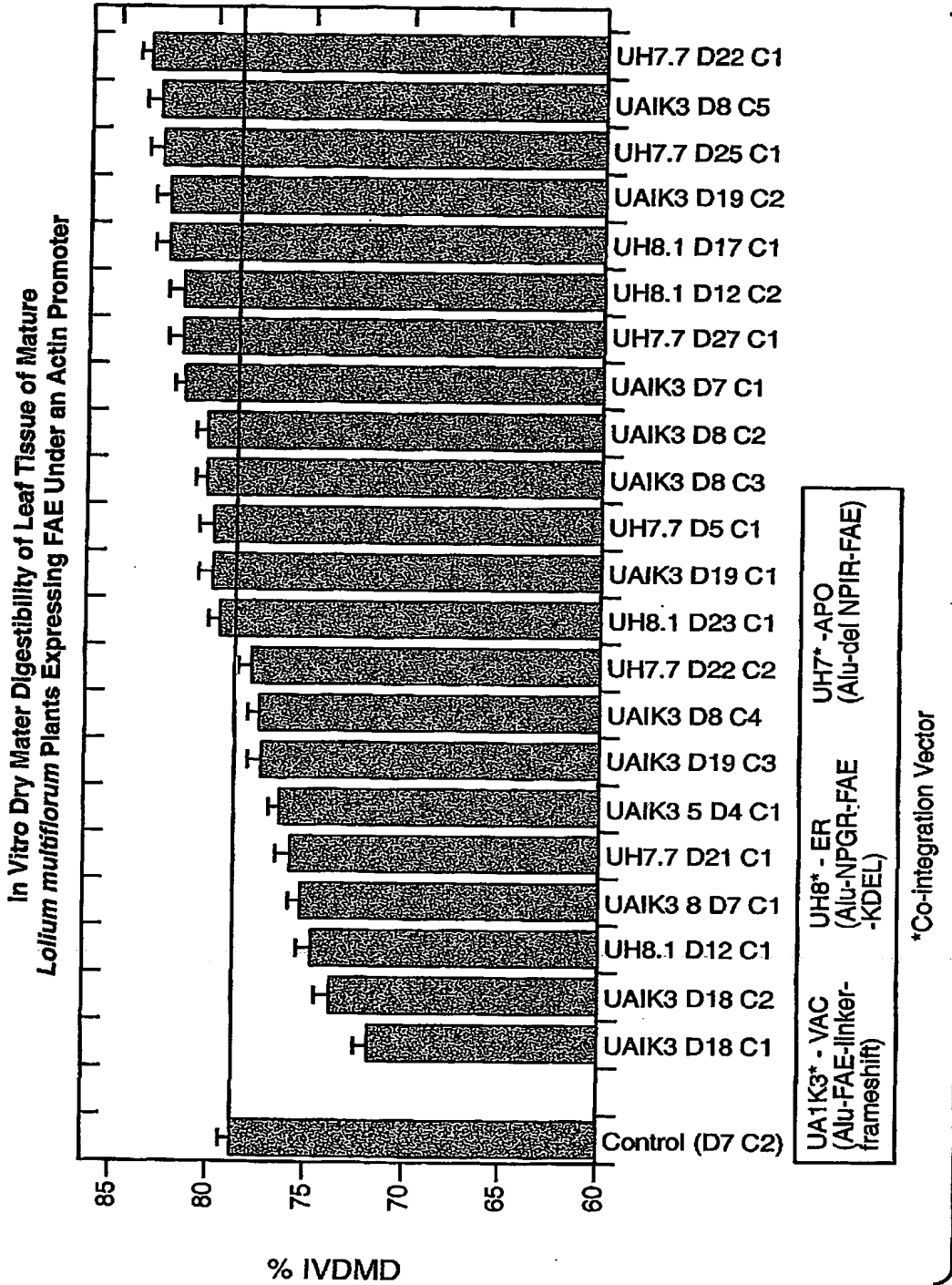


**FIG. 22B**





**FIG.\_23**



**FIG. 24**

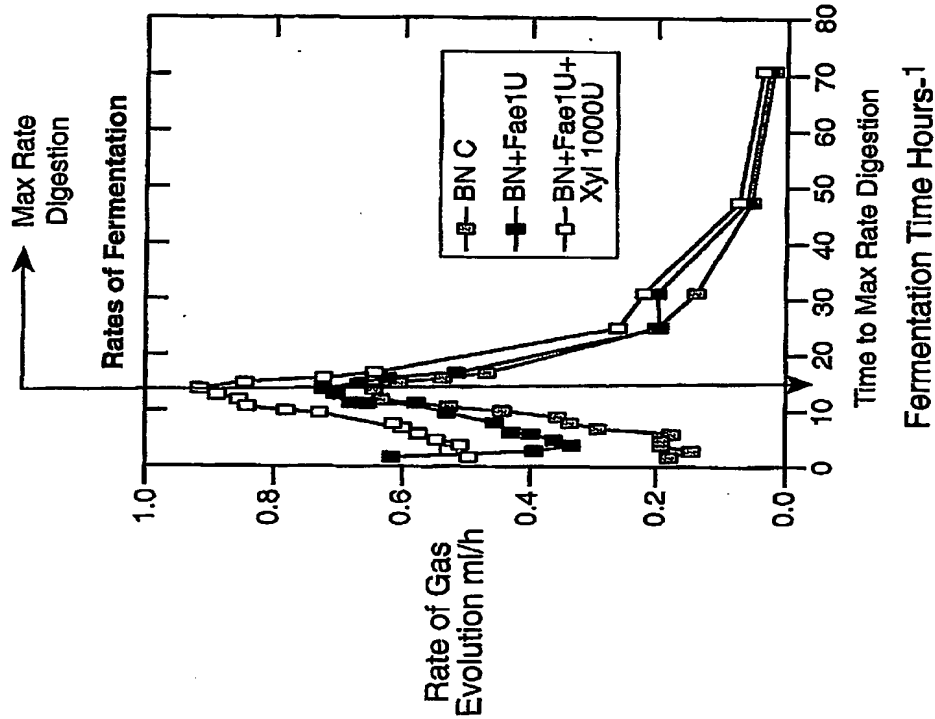


FIG.\_25B

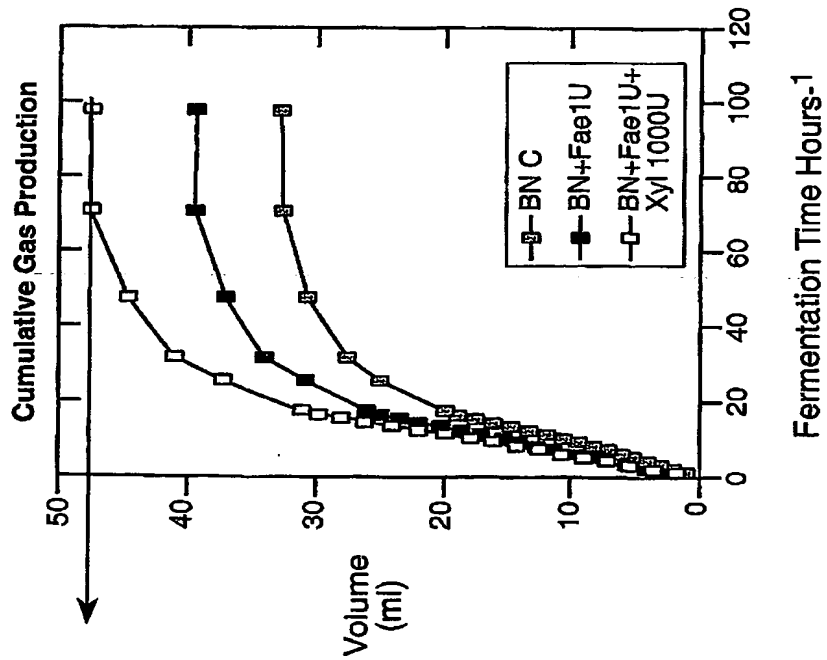
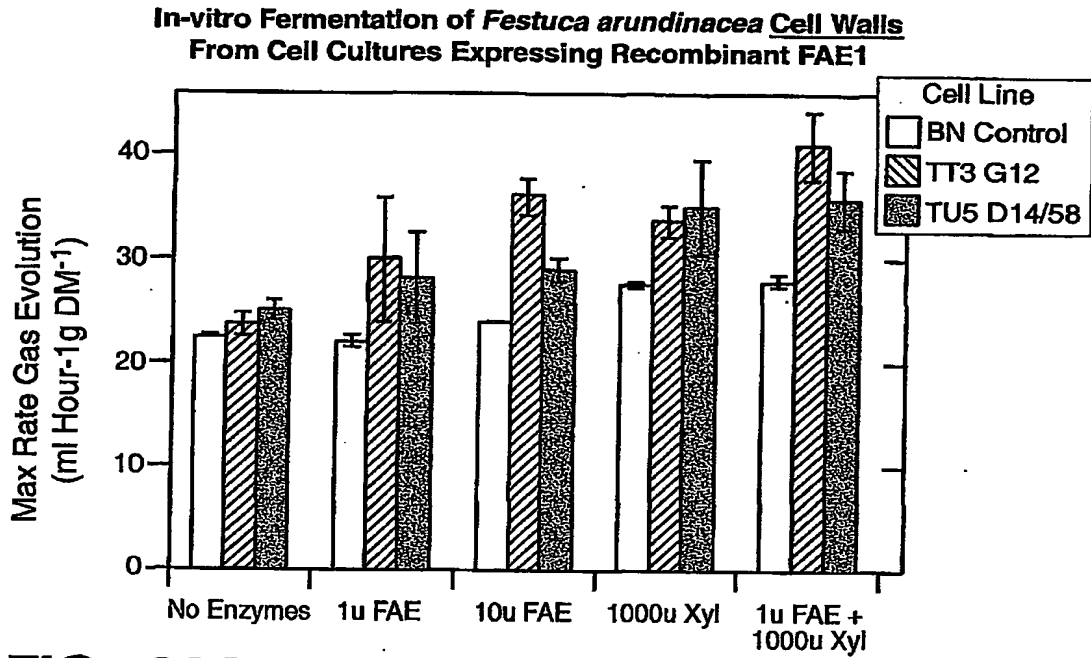
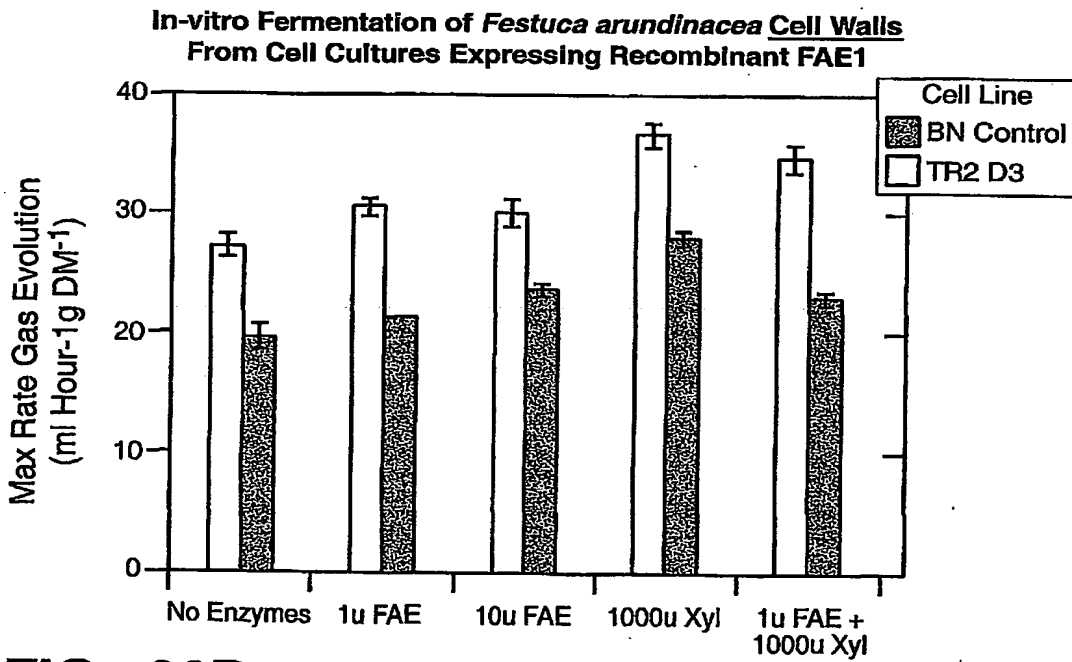


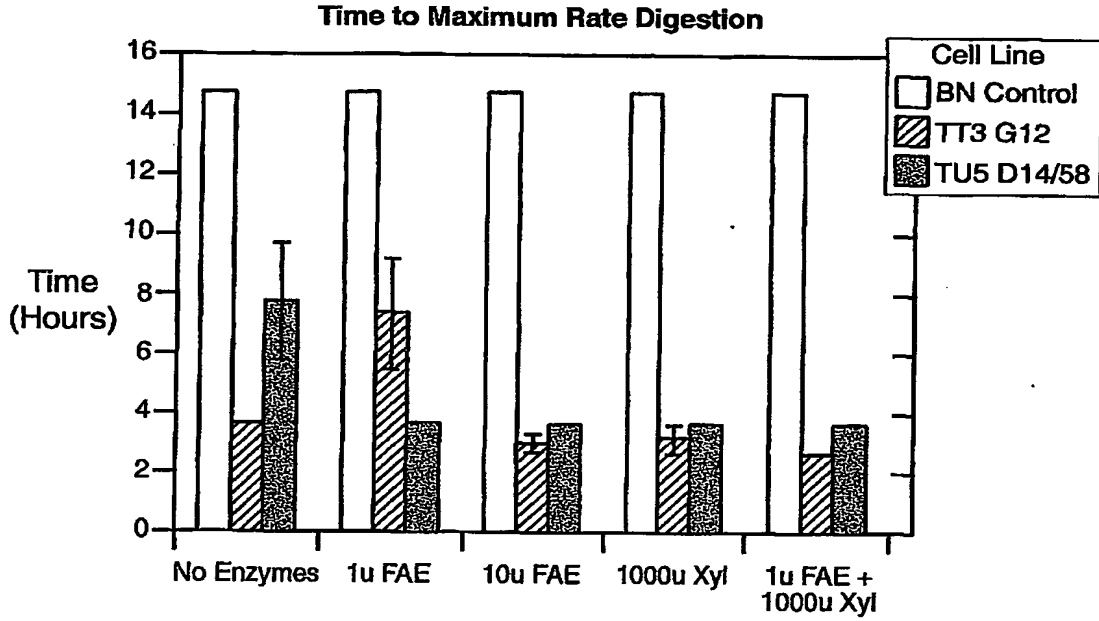
FIG.\_25A



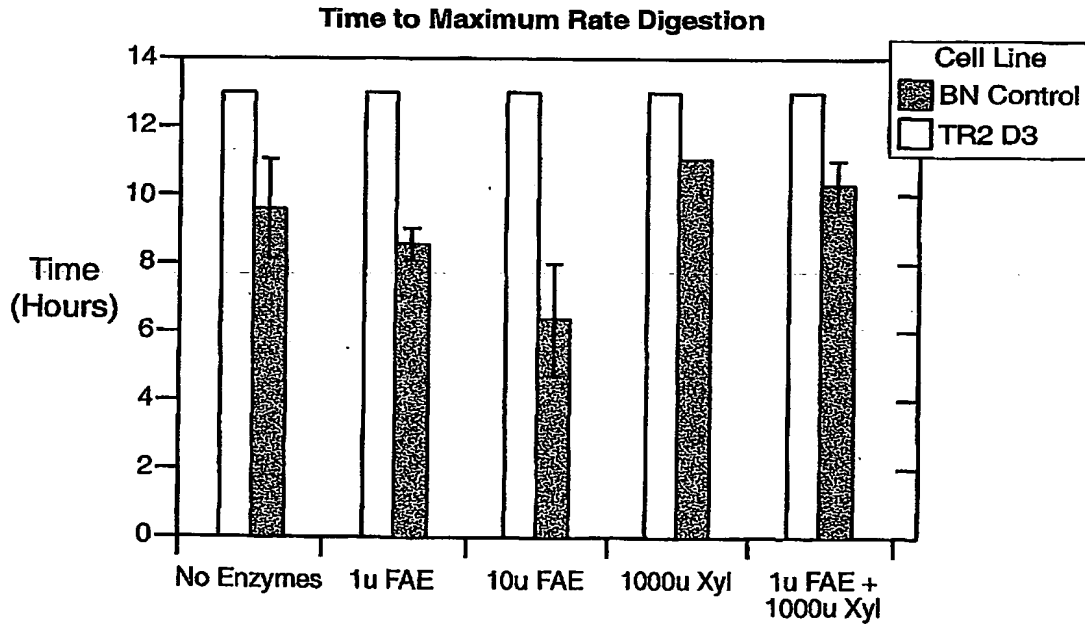
**FIG. 26A** Maximum Rate of Digestion



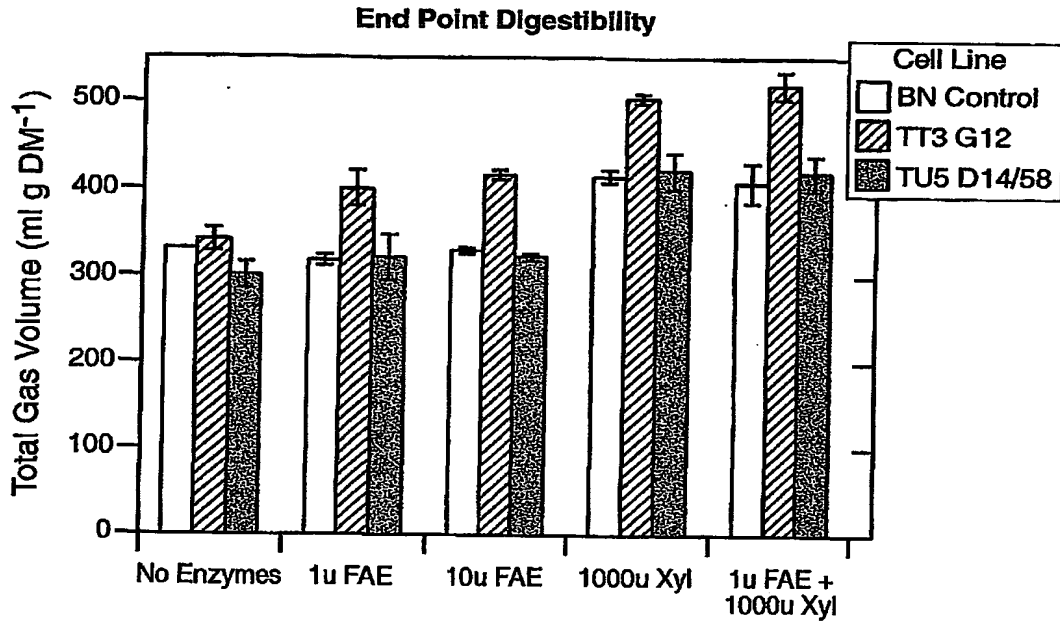
**FIG. 26B** Maximum Rate of Digestion



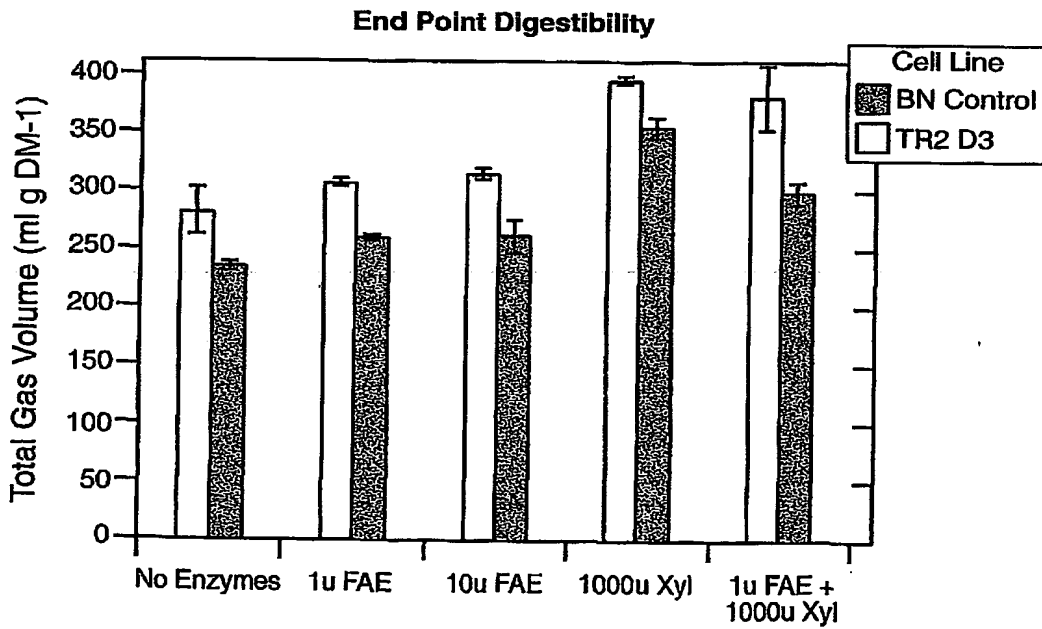
**FIG. 27A**



**FIG. 27B**

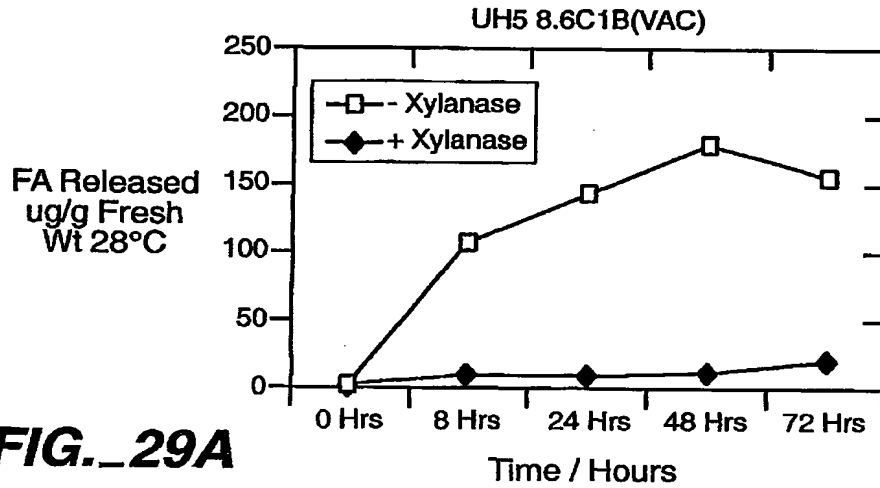


**FIG. 28A**



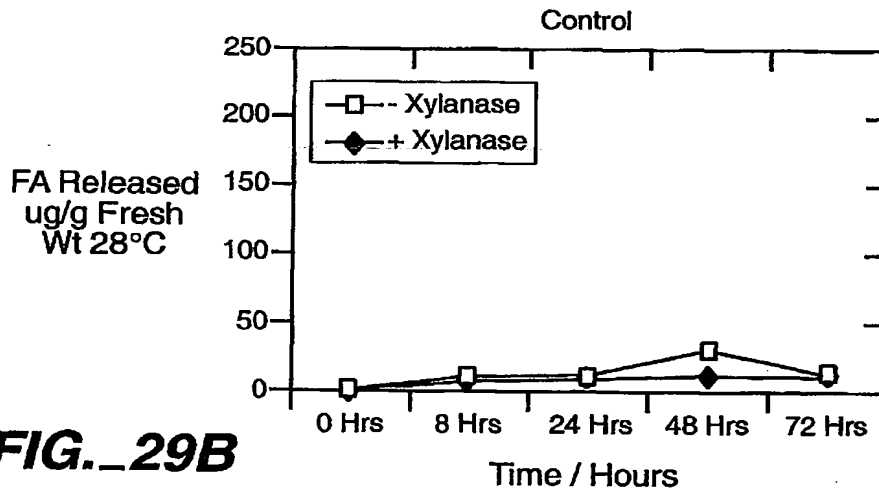
**FIG. 28B**

**Kinetics of FAE Activity by Ferulic Acid Release from Cell Wall under Self Digestion in *Festuca arundinacea* and Stimulation by Xylanase**

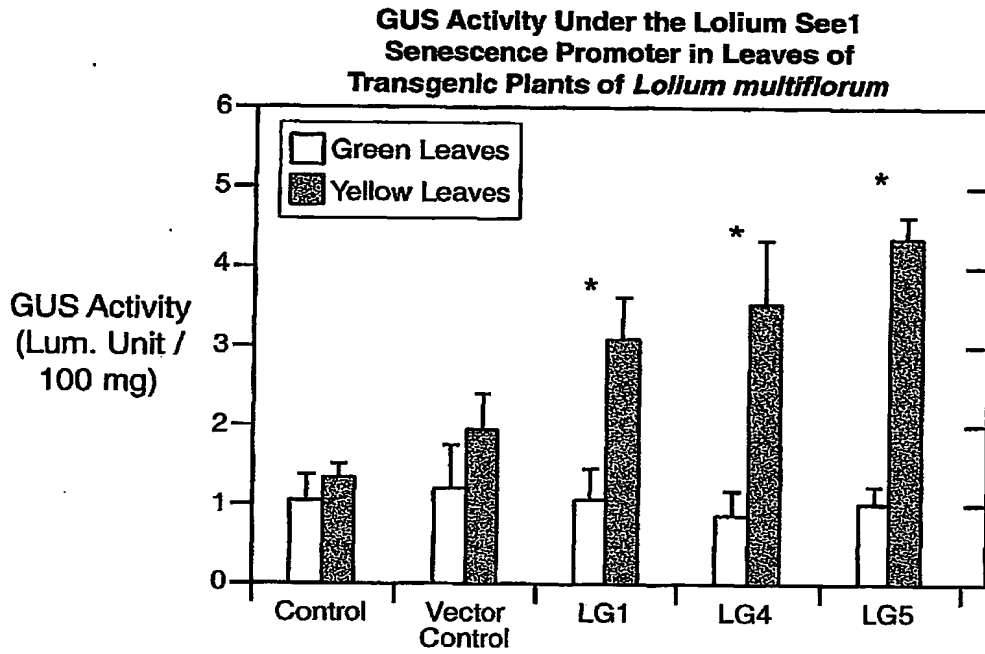


**FIG.\_29A**

**Kinetics of FAE Activity by Ferulic Acid Release from Cell Wall under Self Digestion In *Festuca arundinacea* and Stimulation by Xylanase.**



**FIG.\_29B**



**FIG. 30**



Release of Monomeric and Dimeric HCAs on Self Digestion of Leaves of Vacuolar Targeted FAE Expressing Plants

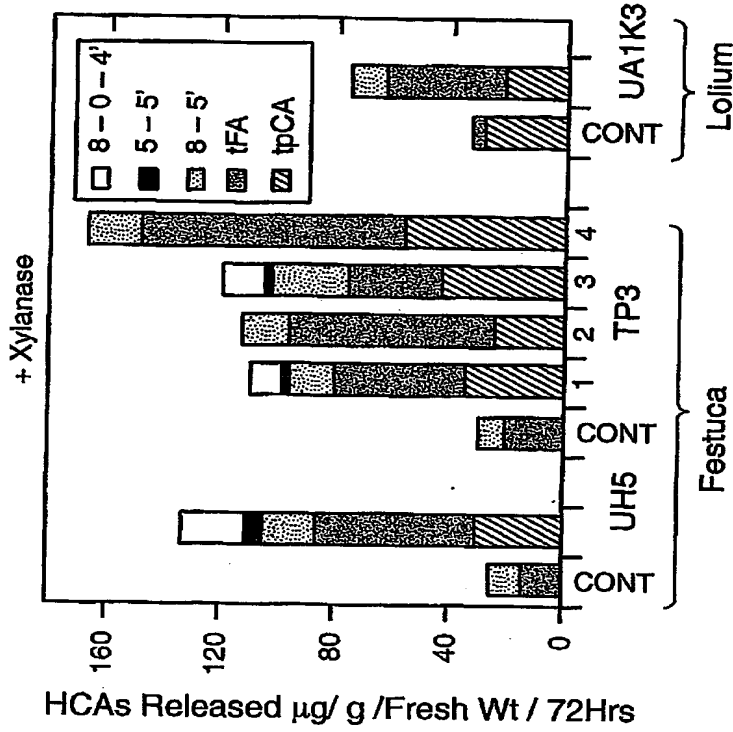


FIG.-31B

Release of Monomeric and Dimeric HCAs on Self Digestion of Leaves of Vacuolar Targeted FAE Expressing Plants

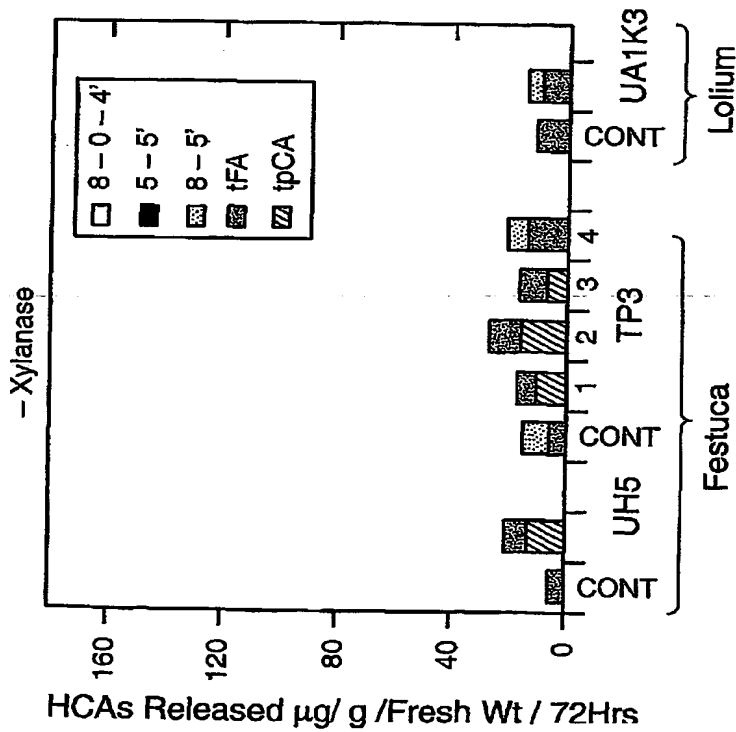
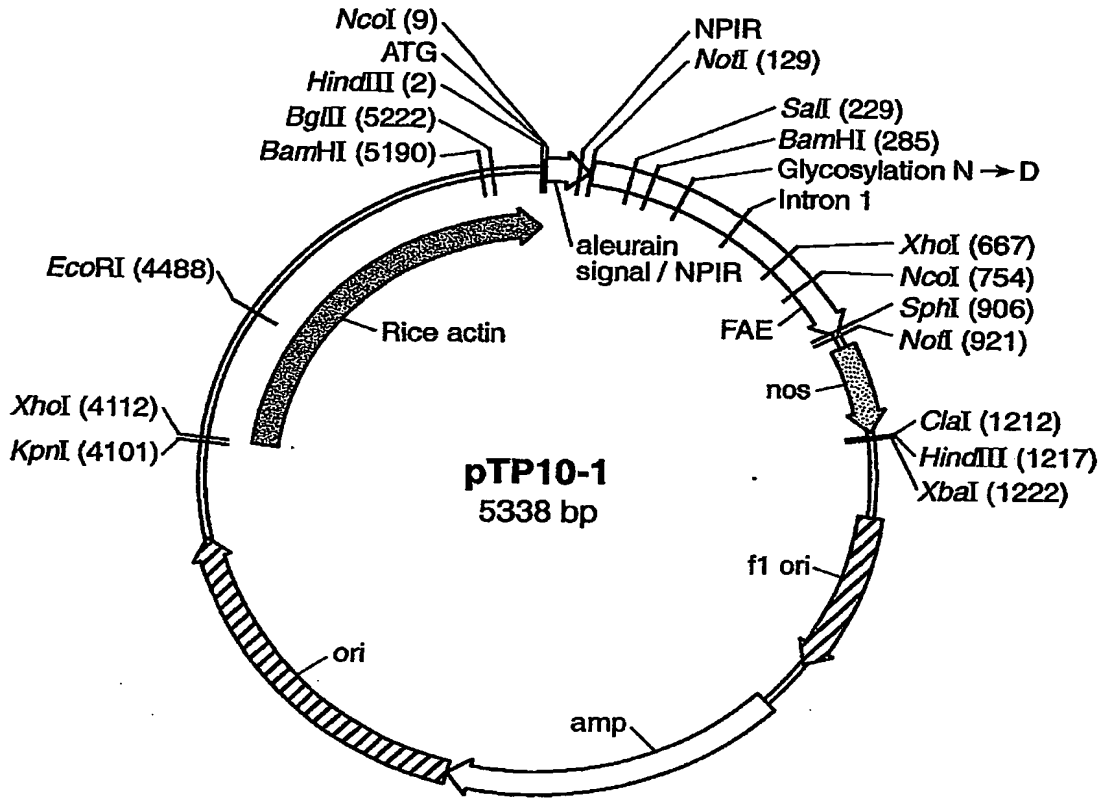


FIG.-31A



**FIG. 32A**

NotI  
\*\*\*\*\*  
HindIII  
\*\*\*\*\*

1 M A H A R V L L L A L A V L A T A A V A V  
AAGCTTACCA TGGCCACGC CCGGTCTTC CTCCTGGGC TCGCCGTGCT GGCACGGCC GCCGTGCGG

NPIR  
\*\*\*\*\*  
NotI  
\*\*\*\*\*

71 . A S S S F A D S N P I R P V T D R A A A S T .  
TGCCTCCTC CTCCTCCTC GCGACTCA ACCGATCCG GCCGTACC GACCGGCGG CCGCTCCAC  
Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAGACC TCTACAGCC TTAGTCGAA ATGCCACTA TCTCCCAAGC TGCCTACGCC

Sali  
\*\*\*\*\*

211 D L C N I P S T I I K G E K I Y N S Q T D I N G  
GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTAACG

BamHI  
\*\*\*\*\*

281 W I L R D D S S K E I I T V F R G T G S D T M .  
GATGGATCCT CCGGACGAC AGCAGCAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA

Glycosylation  
\*\*\*\*\*

351 L Q L D T D Y T L T P F D T L P Q C N G C E V  
TCTACAATC GATACTGACT ACACCTCAC GCCTTCGAC ACCTACCAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
Q Y F D Y A L T V T G H X L G A S L A A L T A .  
491 GCCAGTATCC GGACTACGCG CTGACCTGA CCGGCCACKC CCTCGGGCC TCCCTGGCG CACTCACTGC  
A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCITCGGCG AACCCGCGAG CGGCAATCAG

FIG.-32B

XhoI

631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
GCCTTCGGGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA

NcoI

701 . H A N D G I P N L P P V E Q G Y A H G G V E Y .  
CTCATGCCAA CGACGGCATC CCAACCTGC CCCGGGTGA GCAGGGTAC GCCCATGGG GTGTAGAGTA  
. W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGAGTGCA GTGCTGTGAG

SphI

841 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
GCCAGGGG GACAGGTTGTT GATTAATGCG CACACGACTT ATTTGGGAT GACGAGCGG GCATGCACCT

NotI

KDEI

911 . P V A A A E T T E G \*  
GGCCGGTCCG GCGCGGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTCAA CATTTGGCAA  
981 TAAAGTTTCT TAAGATTGAA TCCTTTGCC GGTCTTGGG TGAATTATCAT ATTAATTCAG TTGAATTCAG  
1051 TTAAGCATGT AATAATTAAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC  
1121 GCAATATAC ATTTAATACG CGATAGAAA CAAATATAG CCGCAAACT AGGATAAAT ATCGCGCGCG

HindIII

ClaI

XbaI

1191 GTGTCACTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCGC CCTATAGTGA  
1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT CGTACTGCG AAACCCCTGG CGTTACCCAA  
1331 CTTAATCGCC TTGCAGCACA TCCCCTTTC GCCAGTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
1401 CTTCCCAACA GTTGGCAGC CTGAAATGGC AATGGGACGC GCCCTGTAGC GCGCAATTA CGCGCGCGCG

FIG.-32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCCTTAC ACTTGCCAGC GCCCTAGCGC CCGTCCCTTT CGCTTTCTTC  
1541 CCTTCTTTTC TCGCCACGTT CCGCCGCTTT CCCCCTAAG CTCTAATCG GGGCTCCCT TTAGGGTCC  
1611 GATTAGTGC TTTACGGCAC CTCGACCCA AAAAATTGA TTAGGGTGAAT GGTTCACGTA GTGGCCCATC  
1681 GCCCTGATAG ACGTFTTTC ACGTFTTTC GCTTGGTCC AGTTCCTTA ATAGTGAAT ATAGTGAAT CTGTGTTCCA  
1751 ACTGGAACA CACTCAACC TATCTGGTC TATCTTTTG TATTTAAGG AATTTGCCG AATTTGCCCT  
1821 ATTGGTTAA AATGAGCTG ATTTAACA AAATTAACG GAATTTAAGG AAAATATTA CGCTTACAAT  
1891 TTAGGTGGCA CTTTTCGGG AAATGCGC AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
1961 TGTATCCGCT CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
2031 TCAACATTTT CGTGTGCCC TTAATCCCTT TTTTGGGCA TTTTGGCTTC CTGTTTTGC TCACCCAGAA  
2101 ACGCTGGTGA AAGTAAAGA TGCYGAAGT CAGTGGGTG CACGAGTGG TTACATCGAA CTGGATCTCA  
2171 ACAGCGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTCTT  
2241 GCTATGTGGC GCGGTATTAT CCGTATTGA CGCCGGGCAA GAGCAACTCG GTCCCGCAT ACACTATTCT  
2311 CAGATGACT TGGTTGAGTA CTCACAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTRAGAGAT  
2381 TATGAGTGC TGCCATAACC ATGAGTATA AACTTCCGGC CAACTTACTT CTGACRAACGA TCGGAGGACC  
2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GACTTCCGC TTGATCGTTG GGAACCGGAG  
2521 CTGAATGAG CCATACCAA CGACGAGCT GACACCAGG TGCCTGTAG TCCCTGTAG AFTGGAAACA ACGTTGCGCA  
2591 AACTATTAC TGGCGAACA CTTACTTAG CTTCCCGCA ACAATTAATA GACTGATGG AGGCGGATAA  
2661 AGTTGCAGGA CCACCTCTG GCTCGCCCT TCCGGCTGG TGGTTTTATG CTGATAAATC TGGAGCCGCT  
2731 GAGCGTGGT CTGCGGYAT CATTCAGCA CTGGGGCAG ATGGHAGCC CTCCGCTATC GTAGTTATCT  
2801 ACACGACGGG GAGTCAGGA ACTATGGAT AACGAATAG ACAGATCGCT GAGATAGGTG CCTCAGTGT  
2871 TAAGCATTGG TAACTGTGAG ACCAAGTTA CTCATATATA CTTTAGATTG ATTTTAAACT TCATTTTTAA  
2941 TTTAARAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCMAAAT CCTTAAACGT GAGTTTTCGT  
3011 TCCACTGAGC GTCAGACCC GTAGAAAAGA TCAAAGGATC TCTTTGAGAT TCTTTGAGAT CCTTTTTC TGCGCGTAA  
3081 CTGCTGCTTG CAACAAAAA ACCACCGCT ACCAGCGGT GTTTGTTTTG CCGATCAAGA GCTACCAA  
3151 CTTTTCCGA AGGTAACCTG CTTCAGCAGA CAATACTGT CAATACTGT CTTCTAGTG TAGCCGTAGT  
3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CTTGCTCTG CTAATCTGT TACCAGTGGC  
3291 TGCTGCCAGT GCGATAAGT CGTCTTAC CCGGTTGGAC TCAAGACGAT AGTTACCAGA TAAGGCGCAG  
3361 CCGTCCGGCT GAACGGGGG TTCGTGCACA CAGCCAGCT TGGAGCGAAC GACCTACACC GAACCTGAGT  
3431 ACCTACAGC TGAGCTATGA GAAAGGCCA CGCTTCCCGA AGGAGAAA GCGGACAGGT ATCCGGTAA  
3501 CGGCAGGTC GMAACAGGAG AGCCACGAG GAGCTTCCA GGGGAAAACG CCTGGTATCT TTATAGTCTT  
3571 GTCGGTTTC GCCACTCTG ACTTGAAGCT CGATTTTGT GATGCTCGT AGGGGGCGG AGCCTATGGA  
3641 AAAACGCCAG CAACCGGCC TTTTACGGT TCCTGGCCCT TTGCTGGCT TTTGCTCACA TGTCTTTTC  
3711 TCGGTTATCC CTTGATCTG TGGATACCG TATTACCGC TTTGAGTGG CTTGATACCG TCGCCGCGC  
3781 CGAACGACCG AGCCAGCGA GTCAGTGA GAGGAGCGG AAGAGCGCC AATACGCAA CCGCTCTCC  
3851 CCGCGCGTTG GCCGATTCAT TAATGCAGT GGCACGACG GTTTCCCGAC TGGAAAAGCGG GCAGTGAGCG

FIG.-32D

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACITTAAGCT TCCGGCTCGT  
 3991 ATGTTGTGTG GAATTTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC  
  
 4061 GCGCAATTAA CCCTCACTAA AGGBAACAAA AGCTGGGTAC CCGGCCCCCC CTCGAGGTCA TTCATATGCT  
 4131 TGAGAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGSTA AGATTACCTG GTCAAAAAGTG AAACATCAG  
 4201 TTAAGAAGTG GTATAAGTAA AATAFCGGTA ATAAAAGGTG GCCCAAAGTG AAATTTACTC TTTTCTACTA  
 4271 TTATAAAAAT TGAGGATGTT TTGTCGGTAC TTGATACCGT CATTTTTGTG TGAATTTGTT TTTAAGTTTA  
 4341 TTCGCCATTT GGAATGCAAT ATCTGTATTT GAGTCGGTGT TTAAGTTCTG TGTCTTTGTA AATACAGAGG  
 4411 GATTTGTATA AGAAATAATCT TTAATAAACCC CATATGCTAA TTTTGACATAA TTTTTCAGAA AAATATATAT  
  
 4481 TCAGGGCAAT TCCACAATGA ACAATAATAA GATTAATAA GCTTGCCCCC GTTGCAGCGA TGGGTATTTT  
 4551 TTCTAGTAAA ATAAAAGATA AACTTAGACT CAATAACATTT ACAAAAACAA CCCCATAAAGT CCTAAAGCCC  
 4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC CACCGCACGT CTCCGAGCCA  
 4691 ACTGGCAAAAT AGTCTCCACC CCGGCACTA TCACCGTGG TTGTCCGGCAG CAGGTGGGTG GGGCCGGAAA  
 4761 AAAAAAAAAG AAGAAGAAA AAAAAGAAA AGAAAACAG CAGGTGGGTG CCGGTCTGCG GGGCCGGAAA  
 4831 AGCGAGGAGG ATCCCGAGCA GCGACGAGGC CCGGCCCTCC CTCCGCTTCC AAGAAAAGC CCCCATCGC  
 4901 CACTATATAC ATACCCCCCC CTCTCTCCC ATCCCCCCTC CACTACACC ACCACACCA CCACCTCCTC  
 4971 CCCCCTCGCT GCCGGACGAC GAGCTCCTCC CCCCCTCCCC TCCGCCGCCG CCGGTAACCA CCCCGCCCTC  
 5041 CTCCTCTTTC TTCTCTCCGT TTTTCTTTCG TCTCGGTCTC GATCTTTGGC CTGTTGAGTT TGGGTGGGCG  
 5111 AGAGCGGCTT CGTCGCCCCAG ATCGGTGCGC GGGAGGGGCG GGATCTCGCG GCTGGCTCT CCGGGCGTGA  
  
 5181 GTCGGCCCGG ATCCTCGCGG GGAATGGGGC TCTCGGATGT AGATCTTCTT TCTTTTCTTCT TTTTGTGGTA  
 5251 GAATTTGAAT CCTCAGCAT TGTTCATCGG TAGTTTCTTCT TTTTCATGAT TGTGACAAAT GCAGCCTCGT  
 5321 GCGGAGCTTT TTTGTAGC

XhoI

KpnI

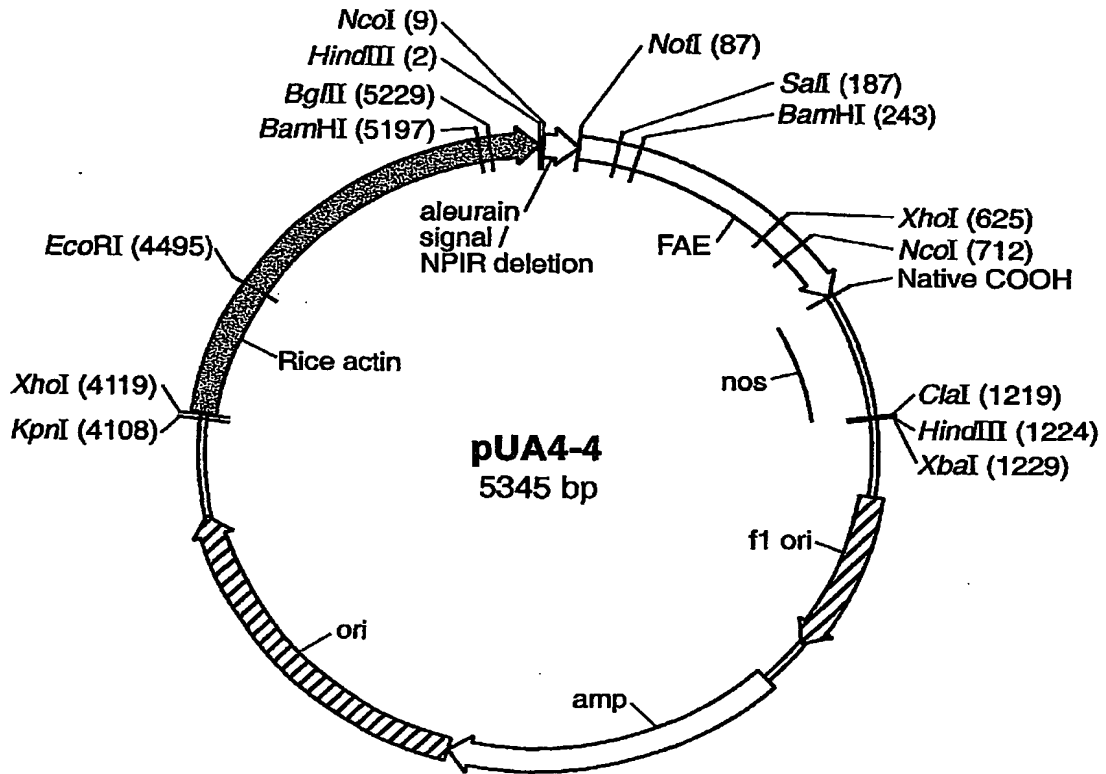
EcoRI

BamHI

BamHI

BglII

FIG.-32E



**FIG. 33A**

NGOI

\*\*\*\*\*

HindIII

\*\*\*\*\*

M A H A R V L L L A L A V L A T A A V A V  
1 AAGCTTACCA TGGCCACAGC CCGGCTCTC CTCTGTCCTC TCGCCGTGCT GGCCACGGCC GCCGTGCGCC

NotI

\*\*\*\*\*

. A S S R A A A S T Q G I S E D L Y S R L V E M .  
71 TCGCCTCCTC CCGGGGGCC GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAT  
Sall

\*\*\*\*\*

. A T I S Q A A Y A D L C N I P S T I I K G E K  
141 GCCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA  
BamHI

\*\*\*\*\*

I Y N S Q T D I N G W I L R D D S S K E I I T V  
211 ATTACAAAT CTCAACTGA CATTAAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATATCACC  
F R G T G S D T N L Q L D T N Y T L T P F D T .  
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAACTAC ACCCTCAGC CTTTCGACAC

. L P Q C N G C E V H G G Y Y I G W V S V Q D Q  
351 CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA  
V E S L V K Q Q V S Q Y P D Y A L T V T G H X L

421 GTCGAGTCGC TTGTCAAACA GCAGGTAGC CAGTATCCGG ACTACGGCT GACCGTACC GGCCACKCCC  
. G A S L A A L T A A Q L S A T Y D N I R L Y T .  
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGGACATAC GACAACATCC GCCTGTACAC  
XhoI

. F G E P R S G N Q A F A S Y M N D A F Q A S S  
561 CTTGGCGGAA CCGGCGAGC GCAATCAGC CTTGGGTCG TACATGAACG ATGCCCTCCA AGCCTCGAGC  
P D T T Q Y F R V T H A N D G I P N L P P V E Q  
631 CCAGATACGA CGCAGTATTT CCGGCTCACT CATGCCAAGC ACGGCATCCC AAACCTGCC CCGGTGGAGC  
NGOI

\*\*\*\*\*

. G Y A H G G V E Y W S V D P Y S A Q N T F V C .

FIG.-33B



701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACA CATTTGTCYG  
 . T G D E V Q C C E A Q G G Q G V N N A H T T Y  
 771 CACTGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAAATGGCA CACGACTTAT  
 F G M T S G A C T W \*  
 841 TTGGGATGA CGAGCGGAGC CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTG CCAGGAAAGA  
 911 TGGATGTCTT GGAGAGGGGG CCGCTAACCC ACTGAAGAT GAGCTGTAAA GAGCAGATC GTTCAAACAT  
 981 TTGGCAATAA AGTTTCTTAA GATTAATCC GATTTCCGGT TTGTCGATGA TTATCATATA ATTTCTGTTG  
 1051 AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA CGTTATTTAT GAGATGGGTT TTTATGATTA  
 1121 GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAACAA AATATAGCGC GCAAACCTAGG ATAAATTATC  
 HindIII

\*\*\*\*\*  
 ClaI XbaI  
 \*\*\*\*\*

1191 GCGCGCGGTG TCATCTATGT TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGTCC AATCGCCCT  
 1261 ATAGTGAGTC GTATTACGG CGCTCACTGG CCGTCTGTTT ACACGCTGCT GACTGGGAAA ACCCTGGCGT  
 1331 TACCCAACCT AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACCC  
 1401 GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT TGCCAGCGCC CTGTAGCGGC GCATTAAGCG  
 1471 CGCGGGGTGT GGTGGTTACG CCGCAGGTTG CCGCTACACT TGCCAGCGCC CTGTAGCGCC CTCTTTTCG  
 1541 TTCTTCCCT TCCTTTCTCG CCACGTTGCG CCGCTTCCCG CGTCAAGCTC TAAATCGGG GCTCCCTTTA  
 1611 GGGTCCGAT TTAGTGTCTT TTAGTGTCTT GACCCCAAAA AACTTGATTA GGGTGTGCT TCACGTAGTG  
 1681 GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT  
 1751 GTTCCAACCT GGAACAACAC TCAACCCCTAT CTCGGTCTAT TCCTTTGATT TATAAGGAT TTTGCCGATT  
 1821 TCGGCCCTAT GGTAAATAA TGAGCTGAT TAAACGCGAA TTTTARCARA ATATTAACCG  
 1891 TTACAATTTA GGTGGCACIT TTCGGGAAA TGTGCGCGGA ACCCTATTT GTTATTTT CTAATATACAT  
 2031 TCMAATATGT ATCCGCTCAT GAGACAATAA CCTGTATAA TGCTTCAATA ATATTAATAA AGGAAAGATA  
 2101 TGAGTATICA ACATTTCCGT GTCCCCCTTA TTCCCTTTTT TCGGGCATTT TGCCTTCCCTG TTTTGTCTCA  
 2171 CCCAGAAACG CTGGTGAAG TAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG  
 2241 GATCTCAACA GCGGTAAAG CTTTGAAGT TTTTCGCCCG AAGAACGTT TCCAATGATG AGCACTTTTA  
 2311 AAGTCTGCT APTGGCGCG GTATATKCC GTATTGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA  
 2381 CTATTTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA  
 2451 AGAGAATTAT GCAGTGTGC CATAACCATG AGTGATAACA CTGGGGCCAA CTTACTTCTG ACAACGATCG  
 2521 GAGGACCGAA GGAGCTAAC GCTTTTTCG ACAACATGGG GGATCATGA ACTCGCCTTG ATCGTTGGGA  
 2591 ACCGGAGCTG AATGAAGCA TACCAACGGA CGAGGTGAC ACCACGATGC CTGTAGCATT GGCACCAACG  
 TATTAACCTG CAACTACTT ACTCTAGCTT CCGGCAACA ATTAATAGAC TGGATGGAGG  
 2661 CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATGCTG ATAAATCTGG

FIG.-33C

2731 AGCCGGTGAG CGTGGGTC TCAGTATCAT GCGCCACTG TGCAGCACTG GGGCCAGATG GFAAGCCCTC CCGTATCGTA  
2801 GTTATCTACA CGACGGGAG TCAGCAACT TCAGGTAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT  
2871 CACTGATTAA GCATTTGGTAA CTGTACAGAC AAGTTTACTC ATATATACTT TAGATGATTT TAAACCTTCA  
2941 TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTAT AATCTCATGA CCAAAATCCC TTAACGTTGAG  
3011 TTTTCGTTCC ACTGAGCGTC AGACCCCGTA ACAAANAAC CACCGTACC AGCGTGGT TGTFTCCGG ATCAAGAGCT  
3081 GCCTAATCTG CTGCTTGCAA ACAAANAAC CACCGTACC AGCGTGGT TGTFTCCGG ATCAAGAGCT  
3151 ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA AFACTGTCTT TCTAGTGTAG  
3221 CCGTAGTTAG GCCACCACCT CAAGAATCTT GTAGCACCGG CTACATACCT CCGCTGTGTA ATCTGTGTAC  
3291 CAGTGGCTGC TGCCACTGGC GATAAGTCTT GTCTTACCAG GTTGGACTCA AGACGATAGT TACCGGATAA  
3361 GGCCAGCGG TCGGGCTGAA CCGGGCTTGT GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA  
3431 CTGAGATACC TACAGCGTGA GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGGTATC  
3501 CCGTAAGCGG CAGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAACGCCCT GGTATCTTTA  
3571 TAGTCTGTG GGGTTTCGCC ACCTTGACT TGAGCGTCTG TTTTGTGTAT GCTCGTCAAG GGGCGGAGC  
3641 CTATGGAAA ACGCCAGCAA CCGGCCCTT TTACGGTCTC TGGCCCTTTT GTCGCTATGT GCTCACATGT  
3711 TCTTCTCTGC GTTATCCCTT GATTCGTGG ATAACCGTAT TACCGCCCTT GAGTGAAGCTG ATACCGCTCG  
3781 CCGCAGCCGA ACGACCAGC GCAGCGAGTC AGTGAAGCGG GAAGCGGAAG AGCGCCAAT ACGCAAACCG  
3851 CCTCTCCCG CCGGTTGGCC GATTCATFAA TGCAGCTGGC ACGACAGGTT TCCCGACTGG AANGCGGBCA  
3921 GTGAGCGCAA CGCAATTAAT GTGAGTTAG TCACTCATTA GGCACCCCGAG GCTTACACT TATGTCTTCC  
3991 GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTTAC

KpnI XhoI

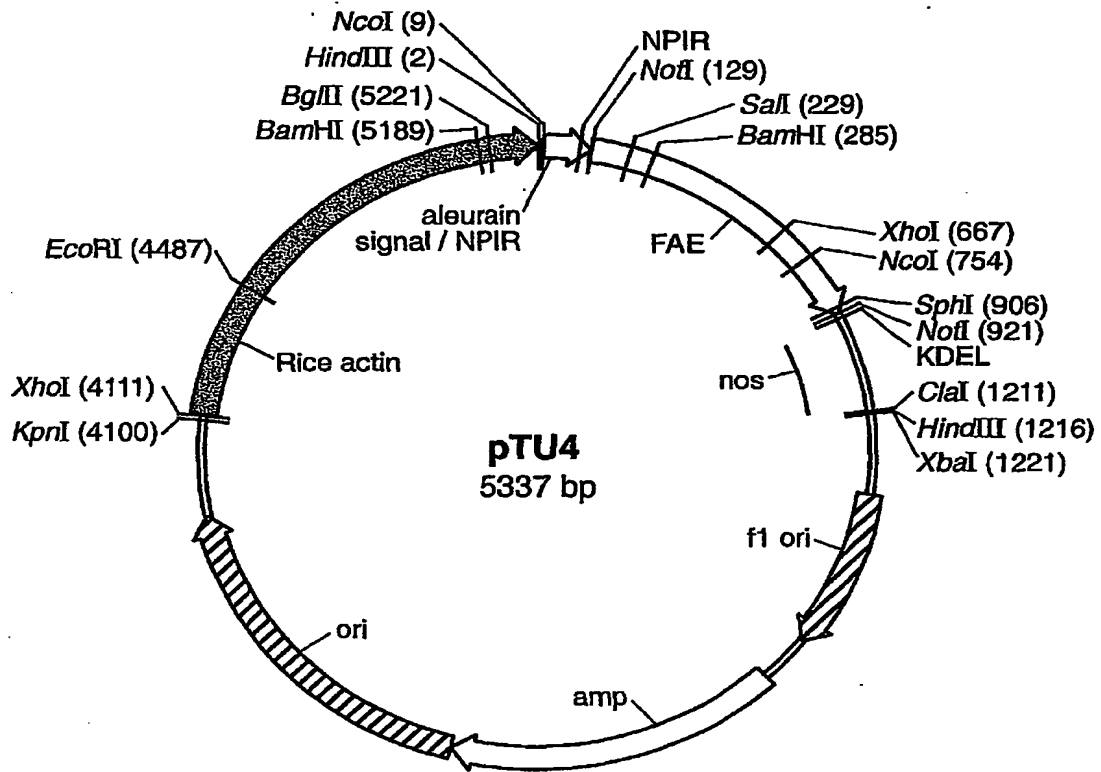
4061 GCCAAGCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCAG GCGCCCCCTC GAGGTCAATC  
4131 ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAC AAAGGTAAAG TTACTTGGTC AAAAGTGAAA  
4201 ACATCAGTTA AAAGGTGGTA TAAATAAAT ATCGGTAATA AAAGGTGGCC CAAAAGTGAAT TTTACTCTTT  
4271 TCTACTATTA TAAAAATTGA GGAFTTTTGG TCGGTACTTT GATACGTCAT TTTTGTATGA ATTGGTTTTT  
4341 AAGTTTATTC GCGATTTTGA AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCTGTGC TTTTGTAAAT  
4411 ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT GACATAAATTT TTGAGAAAAA

EcoRI

4481 TATATATTCA GCGGAATTC ACAAAGAACA ATAATAAGAT TAAAATAGCT TGCCTCCCGTT GCAGCGATGG  
4551 GTATTTTTTC TAGTAAATA AAGATAAAC TTAGACTCAA AACATTTACA AAAACAACCC CTAAGATCCT  
4621 AAAGCCCAA GTGCTATGCA CGATCCATAG CAAGCCAGC CCAACCCMAC CCAACCCAAC CCACCCAGT  
4691 GCAGCCAACT GGCAATAGT CTCCACCCCC GGCATATCA CCGTGAFTG TCCGCACCCAC CGCACGTCCTC  
4761 GCAGCCAAA AAAAAAAG AAAGAAAAA AAAAAAAGA AAAACAGCAG GTGGTCCGG GTCTGTGGGG  
4831 CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAAA GAAACGCCCC

FIG. 33D





**FIG. 34A**

```

      NcoI
      ~~~~~
HindIII
      ~~~~~
1  AAGCTTACCA TGGCCACGC CCGCGTCTC CTCCCTGGCG TCGCCGTGCT GGCACAGCC GCGCTGCGCG
      M A H A R V L L L A L A V L A T A A V A V
      NotI
      ~~~~~
      . A S S S F A D S N P I R P V T D R A A A S T .
71  TCGCCTCCTC CTCCTCCTC GCCGACTCCA ACCGATCCG GCCGTCACC GACCGCGCG CCGCCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
      Sall
      ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAAC T GACATTAACG
      BamHI
      ~~~~~
      . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GAACTAACT ACACCTCAC GCCTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATGG ATGGGTCTCC GTCCAGGACC AAGTCGATC GCTGTCAA CAGCAGGTTA
      . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGG CTGACCGTGA CCGGCCACKC CCTCGGCGC TCCCTGGCGG CACTCACTGC
      . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGGCGAG CGGCAATCAG
      XhoI
      ~~~~~
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCAGATAC GACGCAATAT TTCGCGGTCA
      NcoI
      ~~~~~
      . H A N D G I P N L P P V E Q G Y A H G G V E Y .

```

FIG.-34B

701 CTCATGCCAA CGACGGCATC CCAAAACCTGC CCCCGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCYTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTGA GTGCTGTGAG  
 SphI  
 \*\*\*\*\*

A Q G G Q G V N N A H T T Y F G M T S G A C T W  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGACACTT  
 NotI  
 \*\*\*\*\*

. P V A A A E P L K D E L \*  
 911 GGCCGGTCGC GGCCGGGAA CCACTGAAG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT  
 981 AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT  
 1051 TAAGCATGTA ATAATTAAACA TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG  
 1121 CAATTATACA TTTAATACGC GATAGAANAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG  
 HindIII  
 \*\*\*\*\*

ClaI XbaI

1191 TGTCAATCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGG  
 1261 TCGTATTACG CCGCTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCCAAAC  
 1331 TTAATCGCCT TGCAGCACAT CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCGCA CCGATCGCCC  
 1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG GCGCATTAAG CCGCGCGGGT  
 1471 GTGGTGGTTA CGCGCAGCGT GACCGTACA CTTGCCCAGCG CCCTAGCGCC CGTCCCTTTC GCTTCTTCC  
 1541 CTTCCCTTCT CGCCACGTTT CCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG  
 1611 ATTTAGTGT TTAACGGCAC TCGACCCCAA AAACTTGT TAGGGTGTG GTTCACGTAG TGGGCCATCG  
 1681 CCTGATAGA CGGTTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCCTTAA TAGTGGACTC TTGTTCCAAA  
 1751 CTGGAACAACT ACTCAACCCT ATCTCGGTCT ATTCTTTTGA TTTATAAGG APTTTGCCGA TTTCCGCCTA  
 1821 TTGGTTAAA AATGAGCTGA TTTAACAAA ATTTAACCGG AATTTTAAAC AAATAATAAC GCTTACAATT  
 1891 TAGGTGGCAC TTTTCGGGGA AATGTGCGG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT  
 1961 GTATCCGCTC ATGAGACAAAT AACCTGATA AATGCTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT  
 2031 CAACATTTCC GTGTGCGCCT TATTCCTTCT TTTGCGGCAT TTTGCCCTTC TGTFTTGTCT CACCCAGAAA  
 2101 CGCTGGTGA ATCAAAAGAT GCTGAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA  
 2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCAATGA TGAAGCACTT TAAAGTTCTG  
 2241 CTATGTGGCG CGGTATTATC CCGTATGAC GCGGGCAAG AGCAACTCGG TCGCCGSCATA CACTATTCTC  
 2311 AGAATGACTT GGTGAGTAC TCACCACTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGRAAT  
 2381 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACCTACTTC TGACAACGAT CGGAGGACCG

FIG.-34C

2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCCT TGAATCGTGG GAACCGGAGC  
2521 TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACCAT ACACCCAGGAT GCGTGTAGCA ATGGCAACAA CGTTGCGCAA  
2591 ACTATTAACT GCGGAACATC TTAATCTTAG TTAATCTTAG TTAATCTTAG TTAATCTTAG TTAATCTTAG TTAATCTTAG  
2661 GTTGCAGGAC CACTTCTGCG CCGGCGCTTT CCGGCGCTTT CCGGCGCTTT CCGGCGCTTT CCGGCGCTTT CCGGCGCTTT  
2731 AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGGGCCAGA TGGGGCCAGA TGGGGCCAGA TGGGGCCAGA  
2801 CACGACGGGG AGTCAGGCCA CTATGGATGA ACGAATAGA CAGATCGCTG AGATAGGTGC CTCACATGAT  
2871 AAGCATTTGGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTTAAT  
2941 TTAANAAGGAT CTAGGTGAG ATCTTTTGG ATAAATCTCAT GACCAAAATC CCTTAACGTTG AGTTTTCTGT  
3011 CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAGGATCT CAAGGATCT CAAGGATCT CAAGGATCT CAAGGATCT  
3081 TGTGCTTGC AAACAAMAAA ACCACCGCTA CCAGCGGTGG TTTGTTTTGC CTTTCTAGAG GATFCAAGAG CTACCAACTC  
3151 TTTTTCGGAA GGTAACTGGC TTCAGCAGAG GCCTACATAC CTTCCCTGTC AAATACTGTC TTTCTAGTGT AGCCGTAGTT  
3221 AGGCCACCCAC TTCAAGAACT CTGTAGCAC GCTTCCGAA CAAGACGATA GTTACCCTGGT AACTGAGATA  
3291 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCCTGGT AACTGAGATA  
3361 GGTCCGGCTG AACGGGGGT GAGCTATGAG AAGCGCCAC GCTTCCGAA GGTGTGGACT CAAGACGATA GTTACCCTGGT AACTGAGATA  
3431 CCTACAGCGT GAGCTATGAG AAGCGCCAC GCTTCCGAA GGTGTGGACT CAAGACGATA GTTACCCTGGT AACTGAGATA  
3501 GGCAGGGTCC GAACAGGAGA GCGCAGGAG GAGCTTCCAG GGTGTGGACT CAAGACGATA GTTACCCTGGT AACTGAGATA  
3571 TCGGGTTTCG CCACCTTGA CTTGAGCGTC GATTTTGTG ATGCTCTGCA GGGGGCGGA GCCTATGAA  
3641 AAACGCCACG AACGGGGCT TTTTACGGTT CTTGCCCTTT TGTGGCCCTT TTGCTCACAT GTTCTTTCTT  
3711 GCGTTATCCC CTGATTTCTGT GGAATAACCGT ATTACCGCTT TTGAGTBAE GATFACCGCT CCGCGCAGCC  
3781 GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGAAGCGGA AGAAGCGGA AGAAGCGGA AGAAGCGGA AGAAGCGGA  
3851 CCGCGGTTGG CCGATTCAPT AATGCACTG GCACGACAGG TTTCCCGACT GGAAGCGGG CAGTGAAGCC  
3921 AACGCAATTA ATGAGATTA GCTCACTCAT TAGGCACCC AGGCTTACA CTTTATGCTT CCGGCTCGTA  
3991 TGTGTGTGAG AATTGTGAGC GGAATAACAAT TTCACACAGG AACACGCTAT GACCATGAT ACGCCAAGCG

KpnI XhoI

4061 CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCC TCGAGGTGAT TCATATGCTT  
4131 GAGAAGAGAG TCGGGATAGT CCAAAATAAA ACAAGGTAA GATFACCTGG TCAAAAAGTGA AAACATCAGT  
4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA AATTTACTCT TTTCTACTAT  
4271 TATAAAAAT GAGGATGTTT TGTGGGTACT TTGATACGTC ATTTTTGTAT GAATTTGGTT TTAAGTTTAT  
4341 TCGCGATTTG GAAATGCATA TCTGTATTG AGTCGGTTTT TAAATTCGTT TAAATTCGTT TAAATTCGTT TAAATTCGTT  
4411 ATTTGTATAA GAAATATCTT TAAAANAACC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT

EcoRI

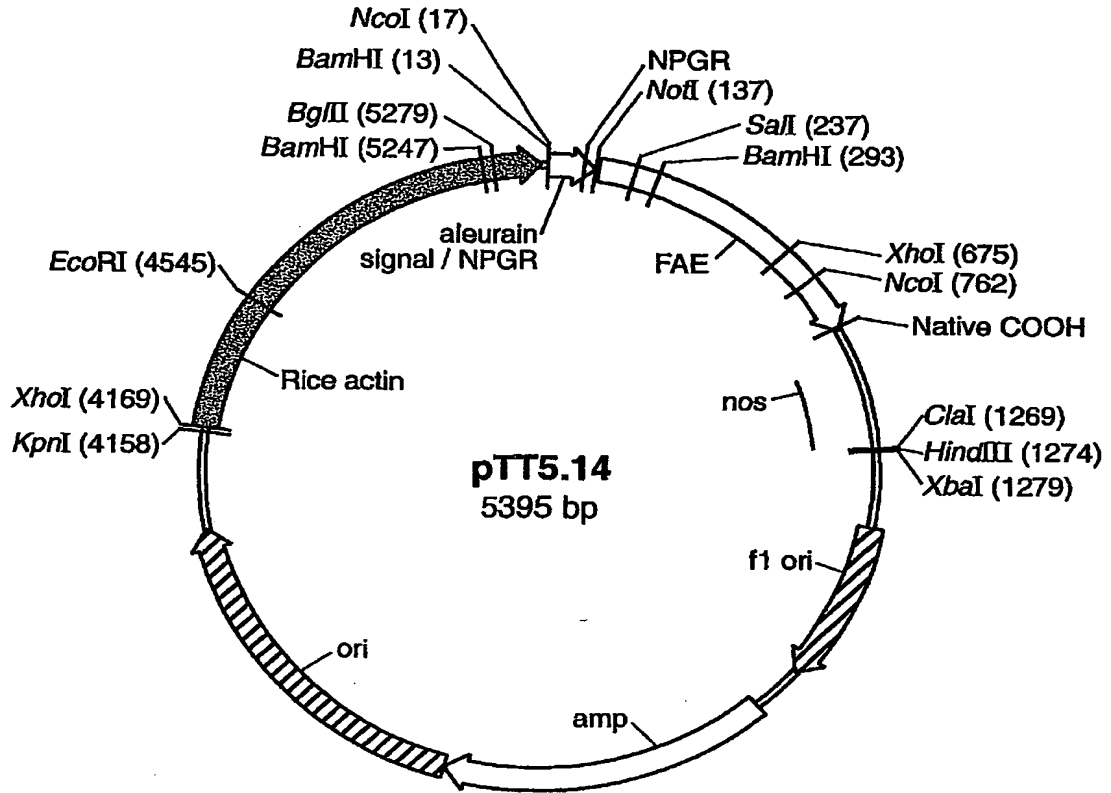
4481 CAGCGCAAT CCACAATGAA CAATAATAAG ATTAATAATG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT  
4551 TCTAGTAAA TAAAAGATA ACTTAGACT AAAACATTA CAAAACAAC CCTTAAAGTC CTTAAAGCCCA

FIG.-34D

4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCACCCCA GTGCAGCCAA  
 4691 CTGGCAATA GTCTCACCC CCGGCACTAT CACCGTGAGT TGTCCGACC ACCGCACGTC TCGCAGCCAA  
 4761 AAAAAAAAAA AGAAGAAA AAAAGAAA GAAAACAGC AGGTGGTCC GGTTCGTGG GGGCCGAAAA  
 4831 GCGAGGAGGA TCGGAGCAG CGACGAGGCC GGGCCCTCCC CCTACCCA CCACCAACC CACCTCCTCC  
 4901 ACTATATACA TACCCCTCCC TCTCTCCCA TCCCTCCCA CCGCCGCGC CGGTAAACC CCGCCCTC  
 4971 CCCCTCGCTG CCGAGCAG AGCTCCCTCC CCTCCCTCC CCGCCGCGC CGGTAAACC CCGCCCTC  
 5041 TCCTCTTCTT TCTCCGTTT TTTTCTCTG CTCGGTCTCG ATCTTTGGC TTGGTAGTTT GGGTGGGCGA  
 5111 GAGCGGCTTC GTGCCCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCCGCG CTGGCGTCTC CGGGCGTGAG  
  
 BamHI  
 \*\*\*\*\*  
 5181 TCGGCCCGGA TCCTCGCGG GAAATGGGCT CTCGGATGTA GATCTCTTTT CTTTCTTCTT TTTGTGGTAG  
 5251 AATTTGAATC CCYCAGCATT GTTCATCGGT AGTTTTTCTT TTTCATGATTT GTGACAAATG CAGCCTCGTG  
 5321 CGGAGCITTT TTGTAGC

FIG.-34E





**FIG. 35A**

```

NcoI
*****
BamHI
*****
1 CCTGACGCCG AGGATCCATG GCCCAGGCC GGTCTCTCCT CCTGGCGCTC GCCGTGCTGG CCACGGCCGC
      M A H A R V L L L A L A V L A T A A .
      NotI
71 CGTCCCGTC GCCTCTCCT CCTCCTTCGC CGACTCCAAC CCGGCCGCG CCGTCACCGA CCGCGCGGCC
      V A V A S S S F A D S N P G R P V T D R A A
      NotI
      ..
141 A S T Q G I S E D L Y S R L V E M A T I S Q A A
      GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT GGCCACTATC TCCCAAGCTG
      Sali
      ..
211 Y A D L C N I P S T I I K G E K I Y N S Q T D .
      CCTACGCCGA CCTGTGCAAC ATTCCGTCCA CTATTATCAA GGGAGAGAAA ATTACAAATT CTCAAACTGA
      BamHI
      ..
281 I N G W I L R D D S S K E I I T V F R G T G S
      CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATATCACC CG TCTTCCGTTG CACTGGTAGT
      D T N L Q L D T N Y T L T P F D T L P Q C N G C
351 GATACGAATC TACACTCGA TACTAATAC ACCCTCAGC CTTTCGACAC CCTACCACAA TGCACGGTT
      E V H G G Y Y I G W V S V Q D Q V E S L V K Q .
421 GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCCG TTGTCAAACA
      Q V S Q Y P D Y A L T V T G H X L G A S L A A
491 GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCC TGGGCGCTC CCTGGCGGCA
      L T A A Q L S A T Y D N I R L Y T F G E P R S G
561 CTCAC TGCCG TGC GACATAC GACAACATCC GCCTGTACAC CTTCGGCGAA CCGCGCAGCG
      XhoI
      ..
631 N Q A F A S Y M N D A F Q A S S P D T T Q Y F .
      GCAATCAGGC CTTCCGCTCG TACATGACG ATGCCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGTATTT
      NcoI
      ..
      R V T H A N D G I P N L P P V E Q G Y A H G G

```

FIG.\_35B

701 CCGGTCACCT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAGC AGGGGTACGC CCATGGCCGGT  
 V E Y W S V D P Y S A Q N T F V C T G D E V Q C  
 771 GTAGAGTACT GGAGCGTTGA TCCATTACAGC GCCCAGAACA CATTGTCTGT CACTGGGGAT GAAGTGCAGT  
 . C E A Q G G Q G V . N N A H T T Y F G M T S G A .  
 841 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATATGCGCA CACGACTTAT TTTGGGATGA CGAGCGGGAGC  
 . C T W \*

911 CTGTACATGG TGATCAGTCA TTTACGCCCTC CCGAGTGA CCAGGAAGA TGGATGTCCT GGAGAGGGGG  
 981 CCGGTAAAC ACTGAAGGAT GAGCTGTAAA GAAGCAGAT GTTCAAACAT TTGGCAATA AGTTTCTTAA  
 1051 GATTGAATCC TGTGCGGGT CTTGCGATGA TTAFCATATA ATTCTGTGTT AATTACGTTA AGCATGTAAAT  
 1121 AATTAAACATG TAAATGATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCCGCA ATTATACATTT  
 1191 TAAATGCGGA TAGAAAACA AATATAGCGC GCAAACCTAGG ATAAATATC GCGCGCGGTG TCATCTATGT

HindIII  
 \*\*\*\*\*  
 \*\*\*\*\*

Clal XbaI  
 \*\*\*\*\*  
 \*\*\*\*\*

1261 TACTAGATCG ATAAAGTTCT AGAGCGGCCG GTGGAGCTCC AATCGCCCT ATAGTGAGTC GTATTACGGC  
 1331 CGCTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCGTGGCGT TACCCAACCT AATCGCCCTG  
 1401 CAGCACATCC CCTTTCGCC AGCTGGCGTA ATAGCGAAGA GCGCCGCACC GATCGCCCTT CCCAACAGTT  
 1471 GCGCAGCCTG AATGGCGAAT GGGACGCGCC CTGTAGCGGC GCATTAAGCG GCGCGGGTGT GGTGGTTACG  
 1541 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGGCCG CTCCTTTCCG TTTCTTTCCCTG  
 1611 CCACGTTCCG CCGCTTTCC CGTCAAGCTC TAAATCGGG GCTCCCTTTA GGGTTCCGAT TTAGTGTCTT  
 1681 ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGG TCACGTAGTG GGCATCGCC CTGATAGACG  
 1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTTCTTAAATA GTGGACTTT GTTCCAAACT GGAACAACAC  
 1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGAT TATAAGGGAT TTTGCCGAT TCGGCCCTATT GGTAAAAAAA  
 1891 TGAGCTGATT TAAACAAAAT TTAACGCGAA TTTTAAACAA ATATTAACGC TTACAATTTA GGTGGCACTT  
 1961 TTCGGGGAAA TGTGCGCGGA ACCCTATTT GTTTATTTTT CTAATACAT TCAATATGT ATCCGCTCAT  
 2031 GAGACAATAA CCTGTATAA TGCCTCAATA ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCGT  
 2101 GTGCGCCCTTA TCCCTTTT TCGGGCATTT TGCCTTCCCTG TTTTTCGCTCA CCCAGAAACG CTGGTGAAG  
 2171 TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAATG GATCTCAACA GCGGTAAGAT  
 2241 CCTTGAGAGT TTTGCGCCCG AAGAACGTTT TCCAATGATG AGCCTTTTA AAGTCTGTCT ATGTGGCGCG  
 2311 GTATTATCCC GTATFGACGC CGGGCAAGAG CAACTCGGTC GCCGATACA CTAFTTCTCAG AATGACTTGG  
 2381 TTGAGTACTC ACCAGTACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGTCTG  
 2451 CATAACCATG AGTGTATAA CTTGCGCCAA CTTACTTCTG ACAACGATG GAGGACCGAA GGAGCTAACCC  
 2521 GCITTTTTTTC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA  
 2591 TACCBAACGA CGAGCGTAC ACCACGATGC CTGTAGCRAT GGCACACAG TTGCGCAAC TATTACTTGG  
 2661 CGAACTACTT ACTCTAGCTT CCCGGCAACA AFTAATAGAC TGGATGGAG CCGATAAAGT TGCAGGACCA

FIG.-35C

2731 CTTCTGGCGT CGGCCCTTCC GGCCTGGCTGG TTTATTGGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC  
2801 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTRAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG  
2871 TCAGGCCAAT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCTT TTTTAAATTT AAAAGGATCT  
2941 CTGTCAAGACC AAGTTTACTC ATATATACTT TAGATTGAT TAAAACTTCA TTTTCTGTTCC ACTGAGCGTC  
3011 AGTGAAGAT CTTTTTGGAT AATCTCANGA CCAAAATCCC TTAACGTGAG TTTTCTGTTCC ACTGAGCGTC  
3081 AGACCCCGTA GAAAGATCA AAGGATCTTC TTGAGATCCC TTTTCTGTTCC GCGTAACTCTG CTGCTTGCAA  
3151 ACAAAAAAAC CACCCCTACC AGCGTGGT TGTGATGCCGG ATCAAGACTT ACCAACTCTT TTTCCGAAGG  
3221 TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT TCTAGTGTAG CCGTAGTTAG GCCACCACCT  
3291 CAAGAACTCT GTAGCACCGC CTACATACTT CGCTCTGCTA ATCTGTATC CAGTGGCTGC TCGCAGTGGC  
3361 GATAAGTCTG GTCTTACCAG GTTGGACTCA AGACGATAGT TACCGGATAA GCGCAGCGG TCGGGCTGAA  
3431 CGGGGGGTTC GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA  
3501 GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGGTATC CGGTAAGCGG CAGGGTCCGA  
3571 ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGTCTCTTAA TAGTCTCTGC GGGTTTCGCC  
3641 ACCCTGACT TGAGCCTCGA TTTTGTGTAT GCTCGTCAAG GGGCGGAGC CTATGGAATA ACGCCAGCAA  
3711 CGCGGCCCTT TTACGGTCTC TGGCCCTTTT CTGGCCCTTT GCTCACATGT TCTTCTCTGC GTTATCCCCT  
3781 GATCTGTGG ATACCGTAT TACCGCCCTT GAGTGAAGCTG ATACCCTCG CCGCAGCCGA ACGATCCGAGC  
3851 GCAGCGAGTC AGTGAAGCGAG GAAAGCGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC  
3921 GATTCATTA TGCAGCTGGC ACAGCAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT  
3991 GTGAGTTAGC TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGTCTGTATG TTGTGTGGRA  
4061 TTGTGAGCGG ATAACTAAT CACACAGGAA ACAGCTATGA CCATGATFAC GCCAAGCGCG CAATTAACCC

KpnI XhoI

4131 TCACATAAGG GAACAAAAGC TGGGTACCAG GCGCCCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC  
4201 GGGATAGTCC AAAATAAAC AAAGGTANGA TTACCTGGTC AAAAGTAAA ACATCAGTTA AARGGTGGTA  
4271 TRAGTAAAAT ATCGTAAATA AAAGGTGGCC CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAAATTGA  
4341 GGATGTTTTG TCGGTACTTT GATACGTCAI TTTTGTATGA ATTTGGTTTT AAGTTATTC GCGATTTGGA  
4411 AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA  

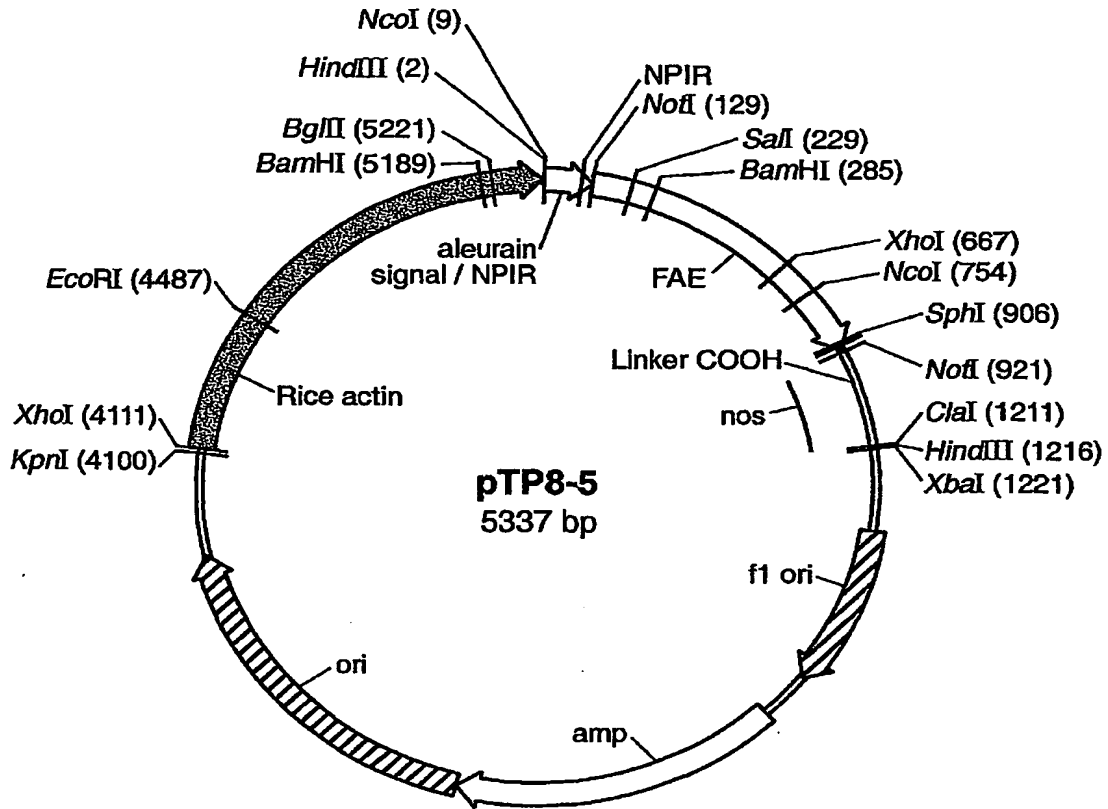
EcoRI

4481 AATATCTTTA AAAACCCAT ATGCTAATTT GACATAATTT TTGAGAAAA TATATATTCA GCGAATTTCC  
4551 ACAATGAACA ATATATAGAT TAAAATAGCT TGCCCCCGTT GCAGCGATGG GTATTTTTTC TAGTAAAAATA  
4621 AAAGATAAAC TTAGACTCAA AACATTACA AAAACAACCC CTAAGTCTCT TAAAGCCAAA GTGCTATGCA  
4691 CGATCCATAG CAAGCCAGC CCAACCCAAC CCAACCCAAC CCACCCCACT GCAGCCAACT GGCAAAATAGT  
4761 CTCCACCCCC GGCACATCA CCGTGAAGTT TCCGCACCAC CCGACGCTC GCAGCCAAA AAAAAAANG  
4831 AAAGAAAAAA AAGAAAAAGA AAAACAGCAG GTGGGTCCCG GTCCGTGGGG CCGGAAAAGC GAGGAGGATC

**FIG. 35D**

4901 GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAAA GAAACGCCCC CCATGCCAC TATATACATA  
 4971 CCCCCCCTC TCCCTCCATC CCCCACACC TACCACCACC ACCACCACCA CCTCCTCCC CCTCGCTGCC  
 5041 GGACGACGAG CTCCTCCCCC CTCCTCCCTC CCGCCGCGG GTAACACACC CGCCCTCTC CTCCTTCTTT  
 5111 CTCCTTTTTT TTTTTCCTCT CCGTCTCGAT CTTTGGCCCTT GGTAGTTTGG GTGGCCGAGA GCGGCTTCGT  
 BamHI  
 5181 CGCCACAGTC GGTGCGCGGG AGGGCCGGA TCTCGCGGCT GGCCTCTCCG GGCCTGAGTC GGCCCGGATC  
 BamHI  
 \*  
 5251 CTCGGCGGGA ATGGGGCTCT CCGATGTAGA TCTTCTTCT TCTTCTTTT TGTGGTAGAA TTTGAATCCC  
 5321 TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA GCCCTCGTCCG GAGCTTTTTT  
 5391 GTAGC

FIG. 35E



**FIG. 36A**

NcoI

HindIII

1 AAGCTTACCA TGGCCACGC CCGGTCTTC CTCTGGCG TCGCCGTGCT GGCACGGCC GCGTGGCCG  
M A H A R V L L A L A V L A T A A V A V

NotI

71 . A S S S F A D S N P I R P V T D R A A A S T .  
TCGCTCCCTC CTCCTCCTC GCCGACTCCA ACCGATCCG GCCGTCACC GACCGCCGG CCGCTCCAC  
. Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC

Sall

211 D L C N I P S F I I K G E K I Y N S Q T D I N G  
GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTTACAA TTCTCAACT GACATTAACG

BamHI

281 . W I L R D D S S K E I I T V F R G T G S D T N .  
GATGGATCCT CCGGACGAC AGCAGCAAG AAATAATCAC CGTCTTCCGT GGCATGGTA GTGATACGAA  
. L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACAATC GATACTACT ACACCTCAC GCCTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATAATGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
. Q Y P D Y A L T V T G H X L G A S L A A L T A .  
491 GCCAGTATCC GGACTACCG CTGACCGTGA CCGGCCACKC CCTCGGGCC TCCCTGGCGG CACTCACTGC  
. A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCGAGCTG TCTGGGACAT ACGACAACAT CCGCTGTAC ACCTTCGGCG AACCGGCAG CGGCAATCAG

XhoI

631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
GCCTTCGGCT CGTACATGAA CGATGCCTC CAAGCCTCGA GCCCAGATAC GACGCAATAT TTCGGGGTCA

NcoI

. H A N D G I P N L P P V E Q G Y A H G G V E Y .

FIG.-36B

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTACA GTGCTGTGAG  
 SphI  
 \*\*\*\*\*

A Q G G Q G V N N A H T T Y F G M T S G A C T W  
 841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCCACCT  
 NotI  
 \*\*\*\*\*

. P V A A A \*  
 \*\*\*\*\*  
 911 GCGCGTCCG GCGCGGTAA CCACTGAAGG ATGAGCTGA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT  
 981 AAAGTTTCTT AAGATTGAAT CCTGTTGCCG GTCTTGGCAT GATTATCATA TAATTTCTGT TGAATPACGT  
 1051 TAAGCATGTA ATAAATAACA TGTAAATGCAT GACGTTAFTT AFGAGATGGG TTTTTATGAT TAGAGTCCCG  
 1121 CAATTATACA TTTAATACGC GATAGAARAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGCG  
 HindIII  
 \*\*\*\*\*

Clal XbaI  
 \*\*\*\*\*

1191 TGTCAATCAT GTTACTAGAT CGATAAGCTT CTAGAGCGCG CGTGGAGCT CCNATTCGCC CTATAGTGAG  
 1261 TCGTATTACG CGCGTCACT GCGCGTGGT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCCCAAC  
 1331 TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
 1401 TTCCAAACAG TTGCGCAGCC TGAATGGCGA ATGGACGCG CCCTGTAGCG GCGCATTAAG CGCGGCGGGT  
 1471 GTGTGGTTA CCGCAGCGT GACCGCTACA CTTGCCAGCG CCTAGAGGCC CGCTCCTTTC GCTTTCTTTC  
 1541 CTTCCCTTCT CGCCACGTTT GCGGCTTTC CCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG  
 1611 ATTTAGTGT TTACGGCACC TCGACCCCAA AAACCTTGTG TAGGGTATG GTTCACGTAG TGGGCCATCG  
 1681 CCCGATAGA CGGTTTTCG CCGTTTGCAG TTGGAGTCCA CGTCTTAA TAGTGGACTC TTGTTCCAAA  
 1751 CTGGAAACAC ACTCAACCTT ATCTCGGTCT ATTTCTTTGA TTTATAAGG ATTTTGGCGA TTTCCGGCCFA  
 1821 TTGGTTAAA AATGAGCTGA TTTAACAATA ATTTAACGCG AATTTAACA AATATAAAC GCTTACAATF  
 1891 TAGGTGGCAC TTTTCGGGA AATGTGCGG AATCCCTAT TTGTTTATTT TTCTAATAAC ATTCMAATAT  
 1961 GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA TAATAATGA AAGAAGAGAG TATGAGTATF  
 2031 CAACATTTCC GTGTCCGCTT TATTCCTTTT TTTGCGGCAT TTTGCTTCC TGTFTTGTCT CACCCAGAAA  
 2101 CGCTGGTGA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA  
 2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCAAATGA TGAGCACTTT TAAAGTTCTG  
 2241 CTATGTGGCG CCGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC  
 2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGMAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATF

FIG.-36C



2381 ATGCAGTGT GGCATAACCA TGAGTGATPAA CACTGCGGCC AACTTACTTC TGACAACGAT TGAGAGGACCG  
 2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC  
 2521 TGAATGAAGC CATAACAAAC GACGAGCCTG ACACCACGAT GCCTGTAGCA ATGGCAACA CGTTGCGCAA  
 2591 ACTATTAACT GCGGAACACT CACTTCTGG CACTTCTGG TCCCGGCCA CAATTAATAG ACTGGATGGA GCGGATPAA  
 2661 GTTGCAGGAC GAGTCTGTC GCGGCTGGT CCGGCTGGT GGTATATGG TGATAAATCT TGAGCCGGTG  
 2731 AGCTGGGTC TCGCGGTATC TCGCGGTATC ATTCAGCAC ATTCAGCAC TGGGCGCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA  
 2801 CACGACGGG AGTCAGGCAA CHTATGATGA CHTATGATGA CAGATCGCTG AGATAGGTGC CTTCACTGATT  
 2871 AAGCAATGGT AACTGTGAGA CCAAGTTTAC TCAATATATAC TTTAGATTGA TTTAAACTTT CATTTTAAAT  
 2941 TTAAGAAGGT CTAGGTGAG ATCTTTTTT ATAAATCTAT GACCAAAATC CTTTAACTGT AGTTTTCGTT  
 3011 CCACTGAGCG TCAGACCCCG TAGAAAGAT CAAGAATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC  
 3081 TGCTGCTTGC AAACAATAA ACCACCGTA CCAGCGGTGG TTTGTTTGGC GGATCAAGAG CTACCAACTC  
 3151 TTTTCCGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC AAATACTGC CTTCTAGTGT AGCCGTAGTT  
 3221 AGGCCACCAC TCAAGAACT CTGTAGCAC CCTACATAC CTCGCTCTGC TAACTCTGTT ACCAGTGGCT  
 3291 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGTGTGGACT CAAGCGATA GTTACCAGAT AAGGCGCAGC  
 3361 GGTCCGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT GGAGCGAAGC ACCTACACCG AACTGAGATA  
 3431 CCTACAGCGT GAGCTATGAG AAGCGCCAC GCTTCCGAA GGGAGAAAGG CCGACAGGTA TCCGGTAAGC  
 3501 GGCAGGCTG GAACAGGAGA GCGCAGGAG GAGCTTCCAG GGGGAAACGC TGGATATCTT TATAGTCTTG  
 3571 TCGGGTTTCG CCACCTCTGA CTTGAGCTG GATTTTGTG ATGCTCGTCA GGGGGCGGA GCTTATGGAA  
 3641 AAACGCCAGC AACGGGGCT TTTTACGGT CTTGAGCTG TGTGGCTT TGTCTCACAT GTTCTTTCTT  
 3711 GCGTTATCCC CTGATCTGT GGAFAACCGT ATTACCGCT TTAGTGAGC TGTATACCGT CCGCCGACCC  
 3781 GAACGACCQA CCGATTCATT AATGCAGTG GCACGACAG AGGAGCCCA ATACGCAAC CGCCTCTCCC  
 3851 CGCGCGTGG CCGATTCATT AATGCAGTG GCACGACAG TTTCCCGACT GGAAGCGGG CAGTGAAGCGC  
 3921 AACGCAATTA ATGTGATTA GCTCACTCAT TAGGCACCCC AGSCTTTACA CTTTATGCTT CCGGCTCGTA  
 3991 TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT ACGCCAAGCG

KpnI

4061 CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCAI TCATATGCTT  
 4131 GAGGAGAGAG TCGGGATAGT CCAAAATPAA ACAAAAGTAA GATTACCTGG TCAAAAAGTGA AAACATCAGT  
 4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA AATTTACTCT TTTCTACTAT  
 4271 TATAAAAATT GAGGATGTTT TGTCCGTACT TTGATACGTC ATTTTGTAT GAAATGGTAT TTAAGTTTAT  
 4341 TCGCGATTGG GAAATGCATA TCTGTATTG AGTCGGTTTT TAAAGTTCGTT GCTTTGTAA ATACAGAGGG  
 4411 ATTTGTATAA GAAATATCTT TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT  
 EGORI

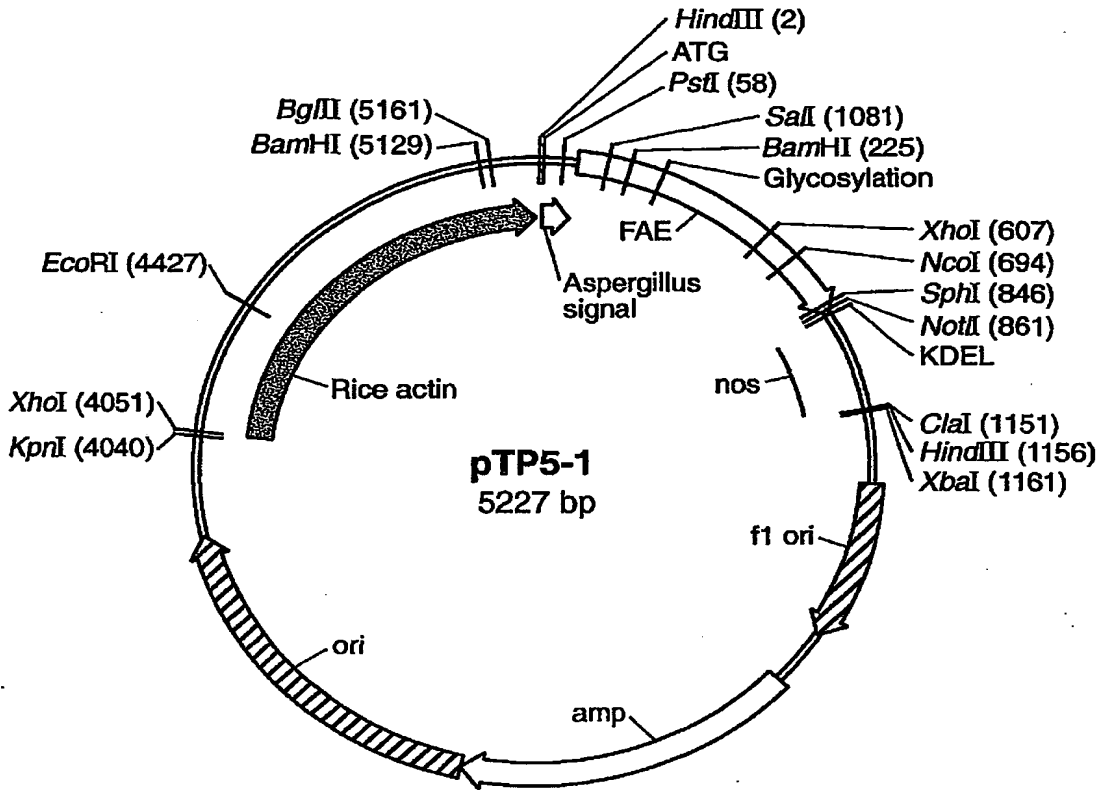
XhoI

4481 CAGCGGAATT CCACAATGAA CAATAATAAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT

FIG. 36D

4551 TCTAGTAAA TAAAGATA ACTTAGACTC AAACATTTA CAAAACAAC CCTAAAGTC CTAAGGCCA  
 4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA GTGCAGCCAA  
 4691 CTGGCAAATA GTCTCCACCC CCGGCACCTAT CACCGTGAGT GAAACACAGC AGGTGGGTCC GGTCTGTGGG GGCCGGAAA  
 4761 AAAAAAAA AGAAGAAA AAAAGAAA AAAAGAAA GAAACACAGC AGGTGGGTCC GGTCTGTGGG GGCCGGAAA  
 4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAAAAACGCC CCCCATCGCC  
 4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CCFACACCA CCACACCCAC CACCTCCTCC  
 4971 CCCCTGGCIG CCGGACGACG AGCTCCTCCC CCTCCCCCCT CCGCCGCCGC CCGTAACCCAC CCCGCCCTC  
 5041 TCCCTCTTCT TCTCCTCTTT TTTTCTTCTG CTCCGCTCCG ATCTTTGGCC TTGGTAGTTT GGGTGGCCGA  
 5111 GAGCGGCTTC GTGCCCAGA TCGGTGCGG GGAGGGCGG GATCTCGCG CTGGCGTCTC CGGCGGTGAG  
BamHI  
~~~~~  
 5181 TCGGCCCGGA TCCTCGCGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTTGTAG  
 5251 AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTTCATGATTT GTGACAAATG CAGCCTCGTG  
 5321 CGGAGCTTTT TTGTAGC

FIG. 36E



**FIG. 37A**

HindIII  
 ~~~~~  
 M K Q F S A K H V L A V V V T A G H A L A  
 1 AAGTTAACA TGAAGCAGTT CTCGCCAAA CAGTCTCTCG CAGTGTGGT GACTGCAGGG CAGCCCTTAG  
 . A S T Q G I S E D L Y S R L V E M A T I S Q A .  
 71 CAGCCTCTAC GCAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC

~~~~~  
 SallI  
 . A Y A D L C N I P S T I I K G E K I Y N S Q T  
 141 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAAC T

~~~~~  
 BamHI  
 D I N G W I L R D D S S K E I I T V F R G T G S  
 211 GACATTAACG GATGGATCCT CCGCGAGCAG AGCAGCAAG AATAATACAC CGTCTTCCGT GGCACCTGGTA  
 . D T N L Q L D T N Y T L T P F D T L P Q C N G .  
 281 GTGATACGAA TCTACAACCT GATACTAACT ACACCTACAC GCCTTTCGAC ACCCTACCAC AATGCAACGG  
 . C E V H G G Y Y I G W V S V Q D Q V E S L V K  
 351 TTGTGAAGTA CACGGTGGAT ATTATATGG ATGGGTCTCC GTCCAGGACC AAGTCGAGT GCTTGTCAAA  
 Q Q V S Q Y P D Y A L T V T G H X L G A S L A A  
 421 CAGCAGTTA GCCAGTATCC GGACTACGCG CTGACCCGTA CCGCCACAC CCTCGGCGCC TCCCTGGCGG  
 . L T A A Q L S A T Y D N I R L Y T F G E P R S .  
 491 CACTCACTGC CGCCAGCTG TCTGGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCCGCCAG

~~~~~  
 XhoI  
 . G N Q A F A S Y M N D A F Q A S S P D T T Q Y  
 561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATAC GACGCCAGTAT

~~~~~  
 NcoI  
 F R V T H A N D G I P N L P P V E Q G Y A H G G  
 631 TTCGGGTCA CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG  
 . V E Y W S V D P Y S A Q N T F V C T G D E V Q .  
 701 GTGTAGAGTA CTGGAGCCTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGGG ATGAAGTGA  
 . C C E A Q G G Q G V N N A H T T Y F G M T S G  
 771 GTGCTGTGAG GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC

FIG.-37B

SphI  
 A C T W P V A A A E P L K D E L \*  
 841 GCATGCACCT GCGCGGTCCG GCGCGGGAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC  
 911 ATTTGGCAAT AAGTTTCTT AAGATTGAAT CCTGTGTCGG GTCTTGGCAT GATTATCATA TAATTTCTGT  
 981 TGAATTACGT TAAGCATGTA AATAATACCA TGTAAATGAT GACGTTATTT ATGAGATGGG TTTTATTATGAT  
 1051 TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAACA AAAATATAGC GCGCAACTA GGATAAATTA

HindIII  
 ClaI XbaI  
 1121 TCGCGCGCGG TGTCACTCAT GTTACTAGAT CGATAAGCTT CTAGAGCGG CCGTGGAGCT CCAATTCGCC  
 1191 CTATAGTGAG TCGTATFACG CCGGCTCACT GGCCGTGCTT TTACAACGTC GTGACTGGGA AAACCCCTGGC  
 1261 GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGNA GAGCCCGCA  
 1331 CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGG CCTGTAGCG GCGCATTAAG  
 1401 CCGCGCGGGT GTGGTGGTTA CCGCAGCCT GACCGCTACA CTTGCCAGCG CCTTAGCGCC CGCTCCTTTC  
 1471 GCTTTCCTCC CTTCCCTTCT CGCCACGCTC CCGGCTTTC TCTAAATCGG TCTAAATCGG GGGTCCCTTT  
 1541 TAGGGTCCG AATTAGTGT TTACGGCAC CCGTCTTTCG CCTTTGACG TTGGAGTCCA CGTCTTTTAA TAGTGGACTC  
 1611 TGGGCCATCG CCTGATFAGA CCGTCTTTCG CCTTTGACG TTGGAGTCCA CGTCTTTTAA TAGTGGACTC  
 1681 TTTGTTCCAA CTGGAACAC ACTCAACCTT ATCTCGGCTT ATTCCTTTCG TTTATAAGGG ATTTTGCCGA  
 1751 TTTTCGGCCTA TTGTTTAAA AATGAGCTGA TTTAACAAA ATTTAACCGG AATTTTAAACA AAATATTAAC  
 1821 GCTTACAATT TAGGTGGCAC TTTTCGGGGA AATGTGGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC  
 1891 ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTGTATA AATGCTTCAA TAATATTGAA AAAGGAGAG  
 1961 TATGAGTATT CAACATTTCC GTGTCCCTTT TATTCCTTTC TTTGCCGGCAT TTTGCCCTTC TGTTTTGTCT  
 2031 CACCCAGAAA CCGTGGTGA AGTAAAGAT GTTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC  
 2101 TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACCT TTTCCAATGA TCGACACTTT  
 2171 TAAAGTTCTG CTATGTGGCG CCGTATATC CCGTATGAC GCCGGCAGG AGCAACTCGG TCGCCGCATA  
 2241 CACTATCTC AGATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGRCAG  
 2311 TAAGAGAAAT ATGCAGTCT' GCCATAACCA TGAATGATA CACTGCGGCC AACTTACTTC TGACAACGAT  
 2381 CCGAGGACCG AAGGAGCTAA CCGTCTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGAATCGTTGG  
 2451 GAACCGGAGC TGAATGAGC CATAACAAAC GACGAGCGTG ACACCCAGAT GCCTGTAGCA ATGGCAACA  
 2521 CGTTGCGCAA ACTATTAAGT GGCGAACATC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA  
 2591 GCGCGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTTG TGAATAATCT  
 2661 GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAAGC ATTGCAAGC TGGTAAAGCC TCCCGTATCG

FIG.-37C

2731 TAGTTATCTA CACGACGGGG AGTCAGGCCA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC  
2801 CTCACGTGAT AAGCATTTGGT AACTGTCAGA CCAAGTTTAC TCAATATATAC TTATAGATTGA TTTAAAACCTT  
2871 CATTITTAAT TTTAAAAGGAT CTAGGTGAAG ATCCTTTTGG TAGAAAAGAT CAAAGGATCT TCTTTGAGATC CTTTTTTTCT  
2941 AGTTTTCGTT CCACTGAGCG TCAGACCCCG TCAGACCCGTA ACCACCCGTA CCAGCGGTGG TTTGTTTGGC GGATCAAGAG  
3011 CGCGTAATC TGCTGCTTGC AATCAAAAAA AATCAAAAAA AATCAAAAAA AATCAAAAAA AATCAAAAAA AATCAAAAAA  
3081 CTACCAACTC TTTTTCGGA GGTAAAGTGC TFCAGCAGAG CCGAGATACC AATCAAAAAA AATCAAAAAA AATCAAAAAA  
3151 AGCCGTAGTT AGCCACCCAC TTCAAAGACT CTGATCAACC GGTGTTGGACT CAAGACGATA GTTACCAGGAT  
3221 ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GTGTGTTGGACT CAAGACGATA GTTACCAGGAT  
3291 AAGGCGCAGC GGTGCGGCTG AACGGGGGTT TCGTGCACAC AGCCAGCTTT GGAGCGAAGC ACCTACACCG  
3361 AACTGAGATA CTTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CCGACAGGTA  
3431 TCCGGTRAAG GGCAGGTCG CAACCTCTGA CTTGAGCGTC GATTTTTTGG ATGCTCGTCA GGGGGGGGGA  
3501 TATAGTCCTG TCGGGTTTCG AAACGCCAGC AACCGGGCCT TTTTACGGTT CCTGGCCCTT TCGTGGCCTT TTGCTCACAT  
3571 GCCTATGGA AACTGATCCC GCGTTATCTG GGATAACCGT ATTACCAGCT TTGAGTGAGC TGATACCGCT  
3641 GTTCTTTCCCT GACTGATCCC GCGTTATCTG GGATAACCGT ATTACCAGCT TTGAGTGAGC TGATACCGCT  
3711 CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCTCA ATACGCAAAC  
3781 CGCCTCTCCC CGCGGTTTGG CCGATTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAGCGGGG  
3851 CAGTGAGCCG AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTTACA CTTTATGCTT  
3921 CCGGCTCGTA TGTGTTGTTGG AATTGTGAGC GGATAACAAT TTCACACACAGG AATCAGCTAT GACCATGATT

KpnI

3991 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGAACAATAA GCTGGGTACC GGGCCCCCCC TCGAGGTGAT  
4061 TCATATGCTT GAGAAGAGAG TCGGGATAGT CCARAATAAA ACRAAGGTAA GATTACCTGG TCAAAAAGTGA  
4131 AATCATCAGT TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTACTCT  
4201 TTTCTACTAT TATAAATTT GAGGATGTTT TGTGCTACT TGTGCTACT TGTGCTACT AATTTGTTAT GAATTTGGTTT  
4271 TTAAGTTTAT TCGCGATTTG GAAATGCATA TCTGTATTTG AGTCGGTTTT TAAGTTCGTT GCTTTTGTAA  
4341 ATACAGAGGG ATTTGTATAA GAATATCTTT TAAAAAACCC ATATGCTAAT TTGACATAT TTTTGAGAAA

EcoRI

4411 AATATATATT CAGCGGATTT CCACAATGAA CAATAAATAG ATTAATAATAG CTTGCCCCCG TTGACGGAT  
4481 GGGTATTTTTT TCTAGTATAA TAAAAGATAA ACTTAGACTC AAAACATTTA CAAAAACAAC CCTTAAAGTC  
4551 CTAAAGCCCA AAGTCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA  
4621 GTGCAGCCCA CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTAGT TGTCCGCAAC ACCGCAAGTC  
4691 TCGCAGCCCA AAAAAAAA AGAAGAAAAA AAAAAACAGC AAAAAACAGC AGGTGGGTCC GGGTCTGTGG

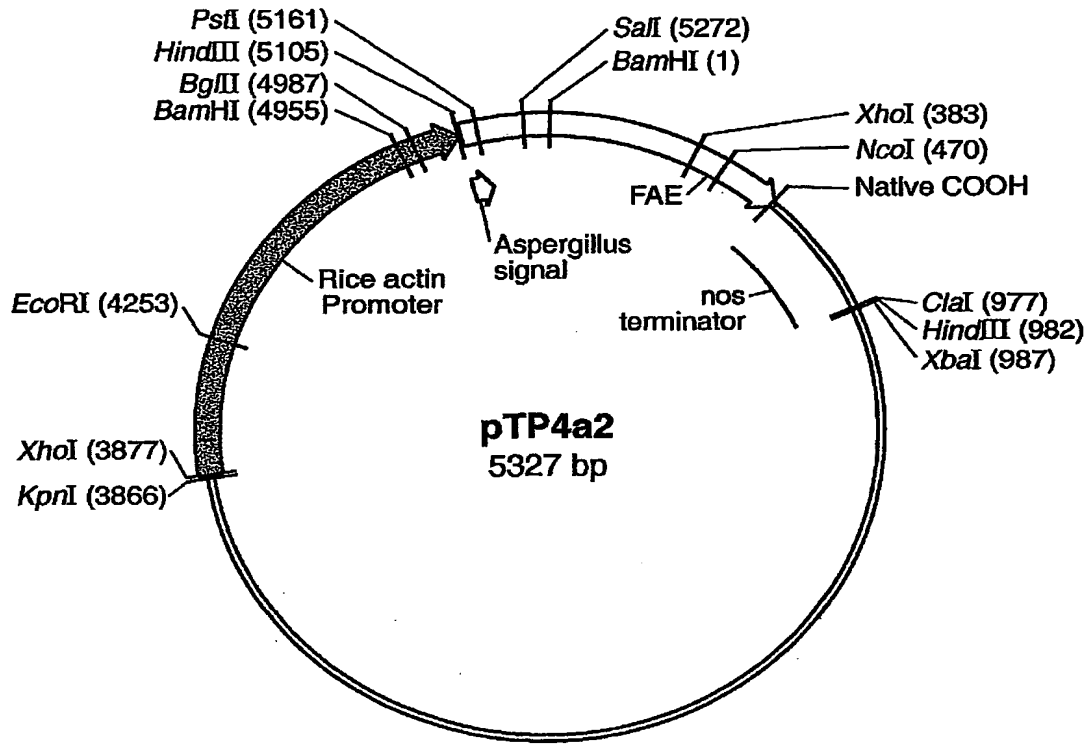
FIG.-37D

```

4761  GSCCGA AAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AGAAACGCC
4831  CCCCATCGCC ACTATATACA TACCCCCCC TCTCCTCCCA TCCCCCAAC CCTACCACCA CCACCACCAC
4901  CACCTCCTCC CCCCTCGCTG CCGACGAGG AGCTCCTCCC CCTCCCCCT CCGCCGCCG CGGTAACCCAC
4971  CCGCCCCCTC TCCTCTTTCT TTCTCCGTTT TTTTTCGT TTTTTCGT CTCGCTCTCG ATCTTTGCC TTGGTAGTTT
5041  GGGTGGCGGA GAGCGGCTTC GTCGCCAGA TCGGTGCGG GGAGGGCGG GATCTCGCG CTGGCGTCTC

      BamHI          BglII
      *****          *****
5111  CCGGCGTGAG TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGAGTGA GATCTCTTT CTTTCTTCTT
5181  TTTGTGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG
5251  CAGCCTCGTG CCGAGCTTTT TTGTAGC
    
```

FIG. 37E



**FIG.\_38A**



BamHI  
 .....  
 . I L R D D S S K E I I T V F R G T G S D T N L  
 1 GATCTCCGC GAGCAGCA GCAAAGAAT AATCACCTC TTCGGTGGCA CTGGTAGTGA TACGAATCTA  
 Q L D T N Y T L T P F D F L P Q C N G C E V H G  
 71 CAATCGATA CTAACATAC CCTCAGCCTTTCGACACC TACCACAATG CAACGGTTGT GAGTACAG  
 . G Y Y I G W V S V Q D Q V E S L V K Q Q V S Q .  
 141 GTGGATATTA TATTGGATGG GTCTCCGTCC AGGACCAATG CGAGTCGCTT GTCAAACAGC AGGTTAGCCA  
 . Y P D Y A L T V T G H X L G A S L A A L T A A  
 211 GTATCCGGAC TACGGCTGA CCGTGACCGG CCACKCCCTC GCGCCTCC TGGCGGCACT CACTGCCGCC  
 Q L S A T Y D N I R L Y T F G E P R S G N Q A F  
 281 CAGCTGCTG CGACATACGA CAACATCCGC CTGTACACCT TCGCGCAACC GCGCAGCGGC AATCAGGCCT  
 XhoI  
 .....  
 . A S Y M N D A F Q A S S P D T T Q Y F R V T H .  
 351 TCGCGTCGTA CATGAACGAT GCCTTCCAA G CCTCGAGCCC AGATACGACG CAGTATTTCC GGGTCACTCA  
 NcoI  
 .....  
 . A N D G I P N L P P V E Q G Y A H G G V E Y W  
 421 TGCCACGAC GGCATCCAA ACCTGCCCC GGTGGAGCAG GGGTACGCC ATGGCGGTGT AGAGTACTGG  
 S V D P Y S A Q N T F V C T G D E V Q C C E A Q  
 491 AGCGTTGATC CTTACAGGC CCAGAACA TTTGTCTGCA CTGGGATGA AGTGCAGTGC TGTGAGGCC  
 . G G Q G V N N A H T T Y F G M T S G A C T W \* .  
 561 AGGCGGACA GGGTGTGAAT AATGCCACA CGACTTATTT TGGGATGACG AGCGGAGCCT GTACATGGTG  
 \*  
 631 ATCAGTCATT TCAGCTCC CGAGTGTACC AGGAAGATG GATGCTCTGG AGAGGGGCC GCGTAACCCAC  
 701 TGAAGGATGA GCTGTAAAGA AGCAGATCGT TCAACATTT GGCATATAAG TTTCTTARA TTGAATCCTG  
 771 TTGCCGGTCT TGCATGATT ATCATATAAT TTCTGTGAA TTACGTTAAG CATGTAATA TTAACATGTA  
 841 ATGCATGACG TTAATTAATGA GATGGGTTTT TATGATTAGA GTCCCGCAAT TATACATTTA ATACCGGATA  
 ClaI  
 911 GAAACAAA TATAGCGGC AAAC TAGGAT AAATTATCGC GCGCGGTGTC ATCTATGTTA CTAGATCGAT  
 XbaI  
 .....  
 HindIII  
 .....

FIG.-38B

981 AAGCTTCTAG AGCGCCGGT GGAGCTCCAA TTCCGCCCTAT AGTGAGTCGT ATTACGGCGG CTCACTGGCC  
 1051 GTCGTTTTAC AACGTCGFGA CTGGGAAAC CTTGGCGTTA CCCACTTAA TCGCCTTGA GCACATCCCC  
 1121 CTTTCGCCAG CTGSCGTAAT AGCGAAGAG CCGGCACCGA ATTAAGCGCG CGCGGTGG CAAGCTGAA  
 1191 TGGCGAATGG GACGGCCCT GTAGCGCGC ATTAAGCGCG CGCGGTGG CTTTCGCGC CAGCGTGACC  
 1261 GCTACACTTG CCAGGCCCT AGCGCCCGT CTTTCGCTT CTTTCGCTT CTTTCGCTT CTTTCGCTT  
 1331 GCTTTCGCCG TCAAGCTTA AATCGGGGG TCCCTTTAGG GTTCCGATTT AGTGTCTTAC GGCACCTCGA  
 1401 CCCCCAAA CTTGATTAG GTGATGTTG CTTTAATAGT GCGACTTGT TCCAAACTGG AACAACACTC  
 1471 TTGACGTTGG AGTCCACGTT CTTTAATAGT GCGACTTGT TCCAAACTGG AACAACACTC  
 1541 CCGTCTATTC TTTTGAATTA TAAGGGAATTT TGCCGATTTG GGCCTATTTG TTAATAAATG AGCTGATTTA  
 1611 ACAAAAATTT AACGGAATTT TTAACAAAAT ATTAACGCTT ACAATTTAGG TGGCACTTTT CCGGGAAATG  
 1681 TCGCGGGAAC CCTATTTGT TTAATTTTCT AAATACATTC AAATATGTA CCGCTCATGA GACAATAACC  
 1751 CTGATAAATG CTTCAATAAT ATTGAAAAG GAAGAGTATG AGTATTC AAC ATTTCCGTTG CGCCCTTATTT  
 1821 CCTTTTITG CGCATTTTGG CCTTCCCTGTT TTTGCTCACC CAGAAACGCT GGTGAAAAGTA AAAGATGCTG  
 1891 AAGATCAGTT GGTGTCACGA GTGGGTACA TCGAACTGGA TCTCAACAGC GGTAAAGATCC TTGAGAGTTT  
 1961 TCGCCCCGAA GAACGTTTTTTC CAATGATGAG CACTTTTAAA GTTTCGCTAT GTGGCGCGGT ATTATCCCCT  
 2031 ATTGACGCCG GGCAGAGCA ACTCGGTCCG CGCATACACT AATCTAGAA TACTTTGTTT GAGTACTCAC  
 2101 CAGTCAACA AAGCAATCTT ACGBATGGA TGACAGTAAG AGAATATAGC AGCTTAACCGG TAACCATGAG  
 2171 TGATAACACT CGGCCAACT TACTTCTGAT AACGATCGGA GGACCGAAGG AGCTAAACCGG TTTTITGAC  
 2241 AACATGGGG ATCATGTAAC TCGCCTTGAT CGTTGGGAAC CGGAGCTGAA TGAAGCCATA CCAACGAGC  
 2311 AGCGTGACAC CACGATGCTT GAGCAATGG CAACAACGTT GCGCAACTA TTAACCTGGC AACTACTTAC  
 2381 TCTAGCTTCC CCGCAACAAT TAAATAGACTG GATGGAGCGG GATAAAGTTG CAGGACCACT TCTCGGCTCG  
 2451 GCCCTTCCGG CTGGCTGGTT TATTTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCGG GGTATCATTG  
 2521 CAGCACITGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG ACGGGGAGTC AGGCAACTAT  
 2591 GGATGACGA AATAGACAGA TCGCTGAGAT AGGTGCCCTCA CTGATTAAGC ATTTGGTAACT GTCAGACCAA  
 2661 GTTTACTCAT ATATACTTTA GATTTGATTA AAACCTCAT TTTTAAATTTAA AAGGATCTAG GTGAAGATCC  
 2731 TTTTITGATA TCTCATGACC AAAATCCCCT AACGTGAGTT TTTCGTTCCAC TGAGCGTACG ACCCCGTAGA  
 2801 AAAGATCAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAA CAAAAAACCA  
 2871 CCGCTACCAG CCGTGGTTTG TTTGCCCCGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA  
 2941 GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGCC CACCACCTCA AGAACTCTGT  
 3011 AGCACCGCCT ACATACCCTG CTCTGCTAAT CCTGTTACCA CCGATAAGG CGCAGCGGTC GGGCTGAACG  
 3081 CTTACCCGGT TGGACTCAAG ACGATAGTTA CCGATAAGG CGCAGCGGTC GGGCTGAACG GGGGTTCGT  
 3151 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAAT GAGATACCTA CAGCGTGAAC TATGAGAAAG  
 3221 CGCCACGCTT CCGAAGGGA GAAAGCGGA CAGGTATCCG GTAAGCGGA GGGTCCGAA CAGGAGCGC  
 3291 ACGAGGAGC TTCCAGGGG AAACGCTGG TATCTTTATA GTCCCTGTCGG GTTTCGCCAC CTCTGACTTG  
 3361 ACGCTCGATT TTTTGTGATG TCGTCAGGGG GCGCGAGCCT ATGGAATAAAC GCCAGCAACG CCGCCTTTTT

FIG. 38C

3431 ACGGTTCCCTG GCCTTTTGGT TCACATGTTT GCCTTTTGGC GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC TATCCCTCGA TTCTGTGGAT  
3501 AACCGTATTA CCGCTTTGA GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC TATCCCTCGA TATCCCTCGA TTCTGTGGAT  
3571 TGAGCCGAGG AGCGGAAGAG CGCCCAATAC GCAACCCGCC TCCTCCCGCG CGTTGGCCGA TTCATTAATG  
3641 CAGCTGGCAC GACAGTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTAGCTC  
3711 ACTCAATTAG CACCCAGGC TTACACATTT ATGCTTCCGG CTGCTAATTT GTGTGGAATT GTGAGCGGAT  
3781 AACAAATTTCA CACAGAAAC AGCTATGATC ATGATTTACGC CAAGCCGCA ATTAACCCCTC ACTAAAGGGA  
  
KpnI XhoI  
3851 ACAAAAGCTG GGTACCGGGC CCCCCTCGA GGTCAATTCAT ATGCTTGAGA AGAGATCGG GATAGTCCAA  
3921 AATAAACAA AGGTAGATT ACCTGGTCAA AGTGAAAC ATCAGTTAAA AGGTGGTATA AGTAAATAT  
3991 CGGTAAATAA AGGTGGCCCA AAGTGAATTT TACTCTTTTC TACTATATA AATATTGAGG ATGTTTTGTC  
4061 GGTAATTTGA TACGTCAATTT TTGTATGAA TGGTTTTTAA GTTTAATTCG GATTTGGAAA TGCATATCTG  
4131 TATTTGAGTC GGTTTTTAAG TTCGTTGCTT TTGTAATATC AGAGGATTT GTATAAGAAA TATCTTTAAA  
  
EcoRI  
4201 AAACCCATAT GCTAATTTGA CATAATTTTT GAGAAAATA TATATTCAGG CGAATTCAC AATGAACAAT  
4271 AATAAGATTA AATAGCTTG CCCCCTGTC AGCGATGGGT ATTTTTCTA GTAAAATAA AGATAAATTT  
4341 AGACTCAAAA CAATACAAA AACAAACCCCT AAGTCTCTAA AGCCCAAGT GCTATGCACG ATCCATAGCA  
4411 AGCCAGCC AACCAACC AACCAACC ACCCCAGTGC AGCCAATGG CAAATAGTCT CCACCCCGG  
4481 CACTATCAC GTGATTTGTC CGCACACC CGGTCCGGGT GGAAGAAGA AAAAAAGAA AGAAAAAAA  
4551 GAAAAAGAAA AACAGCAGGT GGTCCGGGT CAGTCCGGGT GGAAGAAGA GAGGATCGG GAGCAGCGAC  
4621 GAGGCCCGGC CTTCCCTCG CTTCCAAAGA AACGCCCCC ATCGCACTA TATACATACC CCCCCTCTC  
4691 CTCCCATCCC CCAACCTTA CCAACCCAC CACCACCAC TCTCCGCC CCGTGCCTG ACAGCAGCT  
4761 CCTCCCCCTT CCCCCTCGC CGCCGCCGGT AACCAACC CGCCCTCTCT CTTTCTTTCT CCGTTTTTTT  
4831 TTTTCGTCTCG GTCCTGATCT TTGGCCCTGG TAGTTTGGGT GGGCGAGAG GGTTCGTCG CCCAGATCGG  
  
BamHI  
4901 TGCGCGGGAG GGGCGGATC TCGCGGCTGG CGTCTCCGG CGTGAATCG CCCCAGTCTT CGCGGGGAAT  
  
BglII  
4971 GGGGCTCTCG GATGTAGATC TTCCTTTTTC CTTCTTTTGG TGGTAGAAT TGAATCCCTC AGCATTTTTC  
  
HindIII  
5041 ATCGGTAGTT TTTCTTTTCA TGATTTGTGA CAAATGCAG CTCGTGCGGA GCTTTTGTGT AGCAAGCTTA

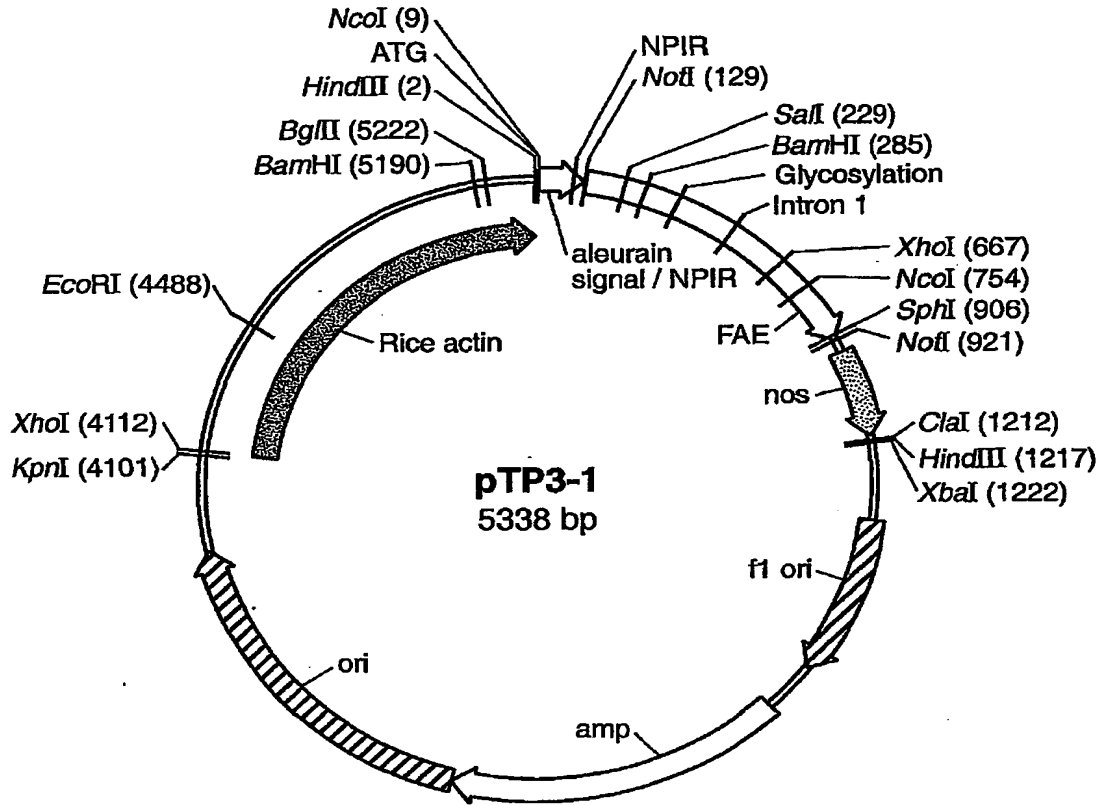
FIG. 38D

```

PstI
*****
M K Q F S A K H V L A V V V T A G H A L A A S .
5111 ACATGAGCA GTTCTCCGCC AARACAGTCC TCGCAGTTGT GGTGACTGCA GGGCAGCCT TAGCAGCCTC
. T Q G I S E D L Y S R L V E M A T I S Q A A Y
5181 TACGCAAGC ATCTCCGAG ACCTCTACAG CCGTTTATGTC GAAATGGCCA CTATCTCCCA AGCTGCCCTAC
SalI
*****
A D L C N I P S T I I K G E K I Y N S Q T D I N
5251 GCCGACCTGT GCAACATTC GTCCACTATT ATCAAGGGAG AGAAATTTA CAATTCFCAA ACTGACATTA
B
. G W
5321 ACGGATG

```

FIG. 38E



**FIG. 39A**

NcoI

\*\*\*\*\*

HindIII

\*\*\*\*\*

1 M A H A R V L L L A L A V L A T A A V A V  
AAGCTTACCA TGGCCACGC CCGCTCCTC CTCCTGGCG TGGCGTGT GGCACGGCC GCGGTGCGC

NPIR

\*\*\*\*\*

NotI

\*\*\*\*\*

. A S S S F A D S N P I R P V T D R A A A S T .  
71 TCGCTCCTC CTCCTCCTC GCGACTCA ACCGATCCG GCGGTACCC GACCGCGCG CCGCTCCAC  
. Q G I S E D L Y S R L V E M A F I S Q A A Y A  
141 GCAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGCCACTA TCTCCCAAGC TGCCATCGCC

Sali

\*\*\*\*\*

D L C N I P S T I I K G E K I Y N S Q T D I N G  
211 GACCTGTCA ACATTCCTGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTAACG

BamHI

\*\*\*\*\*

. W I L R D D S S K E I I T V F R G T G S D T N .  
281 GATGGATCCT CCGGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCACCTGTA GTGATACGAA

Glycosylation

\*\*\*\*

. L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACAATC GATACTAAT ACACCTCAC GCCTTCGAC ACCTACAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT APTAATTTGG ATGGGTCTCC GTCCAGGACC AAGTCAGTC GCTTGTCAA CAGCAGGTTA  
. Q Y P D Y A L T V T G H X L G A S L A A L T A .  
491 GCCAGTATCC GGAATACCG CTGACCGTGA CCGGCCCKC CCTCGGCC TCCCTGGCGG CACTCACTGC  
. A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCAGCTG TCTGGACAT AGCAACAT CCGCTGTAC ACCTTCGGG AACCGGCCAG CGGCAATCAG

FIG.-39B

XhoI

631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
GCCTTCGGGT CFTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATAC GACCCAGTAT TTCCGGGTCA

NcoI

701 . H A N D G I P N L P P V E Q G Y A H G G V E Y .  
CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGCCCAAGAA CACATTTGTC TGCACCTGGG ATGAAAGTGCA GTGCTGTGAG

SphI

841 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT

NotI

KDEL

911 . P V A A A E T T E G \*  
GGCCGGTTCG GCCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA  
981 TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTAATCAT ATAATTTCTG TTGAATTACG  
1051 TTAAGCATGT AATAATTAAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTTATGA TTAGAGTCCC  
1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CCGCAAACT AGGATAATT ATCGCGCGCG

HindIII

ClaI

XbaI

1191 GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC FCCAATTCGC CCTATAGTGA  
1261 GTCGTATTAC GCGCGTCCAC TGGCGTCTGT TTTACAACGT CGTACTGGG AAAACCCCTGG CGTTACCCAA  
1331 CTTAATCGCC TTGCAGCACA TCCCCTTTC GCCAGTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
1401 CTTCCCAACA GTTGGCGAGC CTGAATGGCG AATGGGACCG GCCCTGTAGC GCGGCATTAA GCGCGCGCGG

FIG.-39C

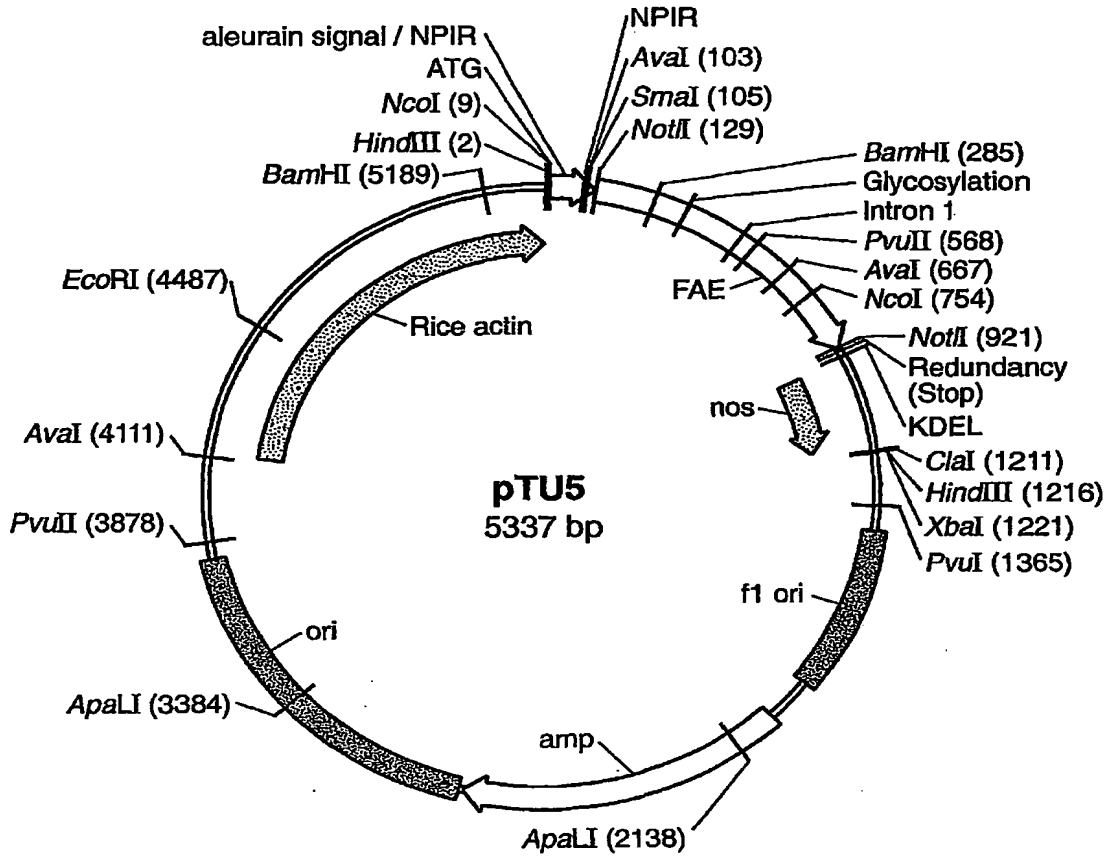
1471 TGTGTGGT ACGGCGAG TGACCGTAC ACTTGGCCAG GCCCTAGCGC CCGCTCCCTT CGCTTCTTC  
1541 CCTTCTTC TCGCCACGTT CGCCGGCTTT CCGCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
1611 GATTTAGTGC TTTACGGCAC CTTGACGGC ACCTGTTTC ACCTGTTTC GATTTAGTGC TTTAGGGTAT GGTTCACGTA GTGGGCCATC  
1681 GCCCTGATAG ACGTCTTTC CTTGACGGC ACCTGTTTC ACCTGTTTC GATTTAGTGC TTTAGGGTAT GGTTCACGTA GTGGGCCATC  
1751 ACTGGAACAA CACTCAACCC CACTCAACCC CACTCAACCC CACTCAACCC CACTCAACCC CACTCAACCC CACTCAACCC  
1821 ATTGGTTAA AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG  
1891 TTAGGTGGCA CTTTTCGGG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG AATGAGCTG  
1961 TGTATCCGCT CATGAGACAA TAACCTGAT TAACCTGAT TAACCTGAT TAACCTGAT TAACCTGAT TAACCTGAT  
2031 TCAACATTTT CCGTCCGCCC CCGTCCGCCC CCGTCCGCCC CCGTCCGCCC CCGTCCGCCC CCGTCCGCCC CCGTCCGCCC  
2101 ACGCTGGTGA AAGTAAAGA AAGTAAAGA AAGTAAAGA AAGTAAAGA AAGTAAAGA AAGTAAAGA AAGTAAAGA  
2171 ACAGCGGTA GATCCTTGAG GATCCTTGAG GATCCTTGAG GATCCTTGAG GATCCTTGAG GATCCTTGAG GATCCTTGAG  
2241 GCTATGTGG CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT  
2311 CAGATGACT TGGTTGAGTA TGGTTGAGTA TGGTTGAGTA TGGTTGAGTA TGGTTGAGTA TGGTTGAGTA TGGTTGAGTA  
2381 TATGCACTGC TGCATATCC TGCATATCC TGCATATCC TGCATATCC TGCATATCC TGCATATCC TGCATATCC  
2451 GAGGAGCTA ACCGCTTTT TGCACAACAT GGGGGATCAT GAACTTACTT CAACTTACTT GAACTTACTT GAACTTACTT  
2521 CTGATGAGG CCAATACAAA CGACGAGCGT GACACCACGA TGCCTGTAGC AATGGCAACA AATGGCAACA AATGGCAACA  
2591 AACTATTAC TGGCGAATA CTTACTCTAG CTTCTCCGCA ACAATTAATA GACTGGATGG AGGGGGATAA  
2661 AGTTGCAGG CCACTTCTGC GCTCGGCCCT TCCGGTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT  
2731 GAGCGTGGT CTTCCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAAGC CTCCCGTATC GTAGTTATCT  
2801 ACACGACGG GATCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACATGAT  
2871 TAAGCATTTG TAACGTGTCAG ACCAAGTTA CTCATATATA CTTTAGANTG ATTTAAACT TCAATTTTAA  
2941 TTTAAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCCTTAACGT GAGTTTTCTG  
3011 TCCACTGAG GTACAGACCC GTAGAAAAGA TCAAAGGATC TTTCTGAGAT CTTTTTTTC TCGCGGTAAT  
3081 CTGCTGCTT CAAACAAA AACCACCGCT ACCAGCGGTG GTTGTGTGC CCGATCAAGA GCTACCAACT  
3151 CTTTTTCCGA AGTAACTGG CTTACAGCAG CCGCAGATAC CAATACTGT CTTCTAGTG TAGCCGTAGT  
3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC GGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
3291 TGTGCCCAGT GGCATTAAGT CGTGTCTTAC CCGGTGGAC TCAAAGCAGT AGTTACCAGG TAAGGGCCAG  
3361 CGGTCCGGCT GATCAGGAGT TTTGTTGAGT TGGAGGAAAC GACCTACACC GAACTGAGAT  
3431 ACCTACAGC TGAAGTATGA GAAAGCGCCA CGCTTCCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAAAG  
3501 CGGACGGTTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGAGGAAACG CCTGGTATCT TTATAGTCTT  
3571 GTCGGGTTTC GCAACGCGCC CACTGAGCTG TTTTACGGT TCGTGGCTT GATGCTGTC AGGGGGCGG AGCCTATGGA  
3641 TCGGTTATCC CCTGATCTG TGGATAACCG TATTACCGC TTTGAGTGA GATGATACCG TCGCCGCGAGC  
3711 CGAACGACCG AGCGCAGCGA GTCAGTGA GAGGAGCGCC AAGACGCAAA AATACGCAAA CCGCCCTCTCC  
3781 CCGCGCGTTG GCCGATTCAT TAATGCAGT GGCACGACAG GTTTCCCGAC TGGAAAAGCGG GCAGTGAAGC  
3851

FIG.-39D



3921 CAACGCAATT AATGTGAGTT AGCTCAGTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT  
 3991 ATGTTGTGTG GAATTTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC  
 XhoI  
 4061 GCGCAATTAA CCTCAGTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTCGAGGTCA TTCAATATGCT  
 4131 TGAGNAGAGA GTCGGGATAG TCCAAAATAA AACAAAGTA AGATTACCTG GTCAAAAGTG AAAACATCAG  
 4201 TTAAAAGGTG GTATAAGTAA AATATCGTAA ATAAAAGGTG GCCCAAAGTG AAATTTACTC TTTTCTACTA  
 4271 TTATAAAAAT TGAGGATGTT TTGTCGGTAC TTGATACGTT CATTTTGTG TGAATTTGTT TTTAAGTTTA  
 4341 TTCGCGATTT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTCTG TGCCTTTGTA AATACAGAGG  
 4411 GATTTGTATA AGAAATATCT TTAATAAAC CATATGCTAA TTTGACATAA TTTTGTAGAA AAATATATAT  
 EcoRI  
 4481 TCAGGCGAAT FCCACAATGA ACAATAATA GATTAATAA GCTTGCCCC GFTGCAGCGA TGGGTATTTT  
 4551 TTCTAGTAAA ATAAAAGATA AACTTAGACT CAATAACATTT ACAAAACAA CCCCTAAAAGT CCTRAAGCCC  
 4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCC CACCGCACGT CTCGCGAGCCA  
 4691 ACTGGCAAAT AGTCTCCACC CCGGCACTA TCACCGTGAG TTGTCCGCAC CAGCGCACGT CTCGCGAGCCA  
 4761 AAAAAAAA AAGAAGAA AAAAAAAA AAAAAACAG CAGGTGGGTG CGGTCTGTGG GGGCCGGAAA  
 4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGC CCGGCCCTCC CTCGGCTTCC AAGAACAAGC CCCCCTCCTC  
 4901 CACTATATAC ATACCCCCCC CTCCTCTCC ATCCCCCAA CCTTACCACC ACCACACCCA CCACCTCCTC  
 4971 CCCCTCGCT GCCGGACGAC GAGTCTCTCC CCCCTCCCC TCCGCCGCCG CCGTAACCA CCCCGCCCCT  
 5041 CTCTCTTTC TTTCTCCGTT TTTTTFTCG TCTCGGTCTC GATCTTTGGC CTTGGTAGTT TGGGTGGGCG  
 5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCGC GGGAGGGCG GATCTCGCG BgIII  
 BstHI  
 5181 GTCGGCCCCG ATCCTCGCG GGAATGGGC TCTCGGATGT AGATCTTCTT TCTTCTTCT TTTTGTGGTA  
 5251 GAATTTGAAT CCTCAGCAT TGTTCAATCG TAGTTTCTT TTTTCAATGAT TGTGACAAAT GCAGCCTCGT  
 5321 GCGGAGCTTT TTTGTAGC

FIG. 39E



**FIG. 40A**

**HindIII NcoI**  
 ~~~~~

1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCCTGGCGC TCGCCGTGCT  
 TTCGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGC AGCGGCACGA

51 GGCCACGGCC GCCGTGCGC TGCCTCCTC CTCCTCCTC GCCGACTCCA  
 CCGGTGCCG CCGCAGCGC AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT

**SmaI**  
 ~~~~~

**AvaI** **NotI**  
 ~~~~~

101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC  
 TGGGCCCGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGAGT CGTCCCCTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC  
 AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCC

201 TGCCTACGCC GACCTGTGCA ACATTCGGTC GACTATTATC AAGGGAGAGA  
 ACGGATGCGG CTGGACACGT TGTAGGCAG CTGATAATAG TTCCTCTCT

**BamHI**  
 ~~~~~

251 AAATTTACAA TTCTCAAAC TACATTAACG GATGGATCCT CCGCGACGAC  
 TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GCGCTGCTG

301 AGCAGCAAAG AAATAATCAC CGTCCTCCGT GGCACTGGTA GTGATACGAA  
 TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAACCTC GATACTAACT ACACCTCCAC GCCTTTCGAC ACCCTACCAC  
 AGATGTTGAG CTATGATTGA TGTGGGAGTG CCGAAAGCTG TGGGATGGTG

401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATFGG ATGGGTCTCC  
 TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451 GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA GCCAGTATCC  
 CAGGTCTCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501 GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG  
 CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC

**PvuII**  
 ~~~~~

551 CACTCACTGC CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC  
 GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG

601 ACCTTCGGCG AACCGCCGAG CCGCAATCAG GCCTTCGCGT CGTACATGAA  
 TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC CCGAAGCGCA GCATGTACTT

**AvaI**  
 ~~~~~

651 CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
 GCTACGGAAG GTTCGGAGCT CGGGTCATG CTGCGTCATA AAGGCCAGT

**FIG. 40B**

SUBSTITUTE SHEET (RULE 26)

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC  
GAGTACGGTT GCTGCCGTAG GGTPTGGACG GGGGCCACCT CGTCCCCTAG

NcoI  
~~~~~

751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCAGAA  
CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CCGGGGTCTT

801 CACATTTGTC TGCCTGGGG ATGAAGTGCA GTGCTGTGAG GCCAGGGCG  
GTGTAAACAG ACGTGACCC TACTTCACGT CACGACACTC CGGGTCCCAG

851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC  
CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCTA CTGCTCGCCG

NotI  
~~~~~

901 GCATGCACCT GGCCGGTCCG GGCCCGGAA CCACTGAAGG ATGAGCTGTA  
CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT  
TTCTTCGTCT AGCAAGTTTG TAAACGTTA TTTCAAAGAA TTCTAACTTA

1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT  
GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAGACA ACTTAATGCA

1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG  
ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

1101 TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC  
AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

1151 AAAATATAGC GCGCAAATA GGATAAATTA TCGCGCGCGG TGTATCTAT  
TTTTATATCG CGCGTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

XbaI  
~~~~~

ClaI HindIII  
~~~~~

1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC  
CAATGATCTA GCTATTCGAA GATCTCGCCG GCCACCTCGA GGTAAAGCGG

1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCTGT TTACAACGTC  
GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT  
CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

PvuII  
~~~~~

1351 CCCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG  
AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC GGGACATCGC

**FIG. 40C**

SUBSTITUTE SHEET (RULE 26)

1451 GCGCATTAAAG CGCGGCGGGT GTGGTGGTFA CGCGCAGCGT GACCCGTACA  
 CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA CPGGCGATGT  
 1501 CFTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCCTCC CTTCTTTCT  
 GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG GAAGGAAAGA  
 1551 CGCCACGTTT CCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT  
 GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTTAGCC CCCGAGGGAA  
 1601 TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA AAACTTGAT  
 ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT TTTTGAACATA  
 1651 TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGTTTTTTCG  
 ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAAGC  
 1701 CCCTTTGACG TTGGAGTCCA CGTCTTTTAA TAGTGGACTC TTGTTCCAAA  
 GGGAAACTGC AACCTCAGGT GCAAGAAAT ATCACCTGAG AACAGGTTT  
 1751 CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCTTTTGA TTTATAAGGG  
 GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAACT AAATATTCCC  
 1801 ATTTTGCCGA TTTCCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA  
 TAAAACGGCT AAAGCCGGAT AACCAATTTT TTA CTGACT AAATTGTTTT  
 1851 ATTTAACGCG AATTTTAACA AAATATTAAC GCTTACAATT TAGGTGGCAC  
 TAAATTGCGC TTAAAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG  
 1901 TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT TTTAAATAC  
 AAAAGCCCCT TTACACGCGC CTTGGGGATA AACAAATAAA AAGATTTATG  
 1951 ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA  
 TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT  
 2001 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCGCCT  
 ATTATAACTT TTTCTTCTC ATACTCATAA GTTGTAAGG CACAGCGGGA  
 2051 TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT CACCCAGAAA  
 ATAAGGGAAA AAACGCCGTA AAACGGAAGG ACAAAAACGA GTGGGTCTTT  

ApaLI  
~~~~~

 2101 CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT  
 GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG TGCTCACCCA  
 2151 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC  
 ATGTAGCTTG ACCTAGAGTT GTCGCCATTC TAGGAACTCT CAAAAGCGGG  
 2201 CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG  
 GCTTCTTGCA AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC  
 2251 CGGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA  
 GCCATAATAG GGCATAACTG CGGCCCGTTC TCGTTGAGCC AGCGGCGTAT  
 2301 CACTATTCTC AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA  
 GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTCGT

**FIG. 40D**

SUBSTITUTE SHEET (RULE 26)

2351 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA  
 AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CGGTATTGGT  
 2401 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG  
 ACTCACTATT GTGACGCCGG TTGAATGAAG ACTGTTGCTA GCCTCTGGC  
 2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT  
 TTCCCTCGATT GGCGAAAAA CGTGTGTAC CCCCTAGTAC ATTGAGCGGA  
 2501 TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG  
 ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC  
 2551 ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAECT  
 TGTGGTGTCTA CCGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA  
 2601 GCGGAECTAC TTAECTTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA  
 CCGCTTGATG AATGAGATCG AAGGGCCGTT GTTAATTATC TGACCTACCT  
 2651 GCGGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT  
 CCGCCTATTT CAACGTCCCTG GTGAAGACGC GAGCCGGGAA GGCCGACCGA  
 2701 GGTTTATTTG TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC  
 CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG  
 2751 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA  
 TAACGTCTGT ACCCCGGTCT ACCATTCCGG AGGGCATAGC ATCAATAGAT  
 2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG  
 GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT GTCTAGCGAC  
 2851 AGATAGGTGC CTCACTGATT AAGCATTGGT AACTGTCAGA CCAAGTTTAC  
 TCTATCCACG GAGTGACTAA TTCGTAACCA TTGACAGTCT GGTTCAAATG  
 2901 TCATATATAC TTTAGATTGA TTTAAAACTT CATTTTTAAT TTTAAAAGGAT  
 AGTATATATG AAATCTAECT AAATTTTGAA GTAAAAATTA AATTTTCCTA  
 2951 CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAAATC CCTTACGTTG  
 GATCCACTTC TAGGAAAAAC TATTAGAGTA CTGGTTTTAG GGAATFGCAC  
 3001 AGTTTTCTGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
 TCAAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA  
 3051 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAA  
 AGAACTCTAG GAAAAAAGA CCGCATTAG ACGACGAACG TTTGTTTTTT  
 3101 ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC  
 TGGTGGCGAT GGTCCGCCACC AAACAAACGG CCTAGTTCTC GATGGTTGAG  
 3151 TTTTTCCGAA GGTAECTGGC TTCAGCAGAG CGCAGATACC AAATACTGTC  
 AAAAAGGCTT CCATFGACCG AAGTCGTCTC GCGTCTATGG TTTATGACAG  
 3201 CTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC  
 GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCTTGA GACATCGTGG  
 3251 GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG  
 CGGATGTATG GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCACT

**FIG. 40E**

SUBSTITUTE SHEET (RULE 26)

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT  
CGCTATTTCAG CACAGAAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

## ApaLI

~~~~~

3351 AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT  
TTCCGCGTCCG CCAGCCCAGC TTGCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
CCTCGCTTTCG TGGATGTGGC TTGACTCTAT GGATGTCGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC  
TTTCGCGGTG CGAAGGGCTT CCTCTTTCC GCCTGTCCAT AGGCCATTCC

3501 GGCAGGGTCCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC  
CCGTCACCAGC CTTGTCTCTC CGCGTGCTCC CTCGAAGGTC CCCCTTTGCG

3551 CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC  
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG

3601 GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC  
CTAAAACAC TACGAGCAGT CCCCCGCCT CGGATACCTT TTTGCGGTCC

3651 AACGCGGCCCT TTTTACGGTT CTTGGCCTTT TGCTGGCCTT TTGCTCACAT  
TTGCGCCGGA AAAATGCCAA GGACCGGAAA ACGACCGGAA AACGAGTGTA

3701 GTTCTTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT  
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG  
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CGCGTCGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC  
AGTCACTCGC TCCTTCGCCT TCTCGCGGGT TATGCGTTTG GCGGAGAGGG

## PvuII

~~~~~

3851 CGCGCGTTGG CCGATTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT  
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT  
CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG  
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT  
TTAACACTCG CCTATTGTTA AAGTGTGTCC TTTGTCGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC  
TGCGGTTCGC GCGTTAATTG GGAGTGATTT CCTTGTTTT CGACCCATGG

## AvaI

~~~~~

4101 GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT  
CCCAGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

**FIG. 40F**

SUBSTITUTE SHEET (RULE 26)

4151 CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAGTGA AAACATCAGT  
 GGTTTTATTT TGTTTCCATT CTAATGGACC AGTTTTCACT TTTGTAGTCA

4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA  
 ATTTTCCACC ATATTCATTT TATAGCCATT ATTTTCCACC GGGTTTCACT

4251 AATTTACTCT TTTCTACTAT TATAAAAATT GAGGATGTTT TGTCCGTACT  
 TTAAATGAGA AAAGATGATA ATATTTTAA CTCTACAAA ACAGCCATGA

4301 TTGATACGTC ATTTTTGTAT GAATGGTTTT TTAAGTTTAT TCGCGATTTG  
 AACTATGCAG TAAAAACATA CTTAACCAA AATTCAAATA AGCGCTAAAC

4351 GAAATGCATA TCTGTATTTG AGTCGGTTTT TAAGTTCGTT GCTTTTGTAA  
 CTTTACGTAT AGACATAAAC TCAGCCAAA ATTCAGCAA CGAAAACATT

4401 ATACAGAGGG ATTTGTATA GAAATATCTT TAAAAAACC ATATGCTAAT  
 TATGTCFCCC TAAACATATT CTTTATAGAA ATTTTTTGGG TATACGATTA

EcoRI  
 ~~~~~

4451 TTGACATAAT TTTTGAGAAA AATATATATT CAGGCGAATT CCACAATGAA  
 AACTGTATTA AAAACTCTTT TTATATATAA GTCCGCTTAA GGTGTTACTT

4501 CAATAATAAG ATTAAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT  
 GTTATTATTC TAATTTTATC GAACGGGGGC AACGTCGCTA CCCATAAAAA

4551 TCTAGTAAAA TAAAAGATA ACTTAGACTC AAAACATTTA CAAAAACAAC  
 AGATCATTTT ATTTTCTATT TGAATCTGAG TTTTGTAAAT GTTTTTGTG

4601 CCCTAAAGTC CTAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA  
 GGGATPTCAG GATTCGGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT

4651 GCCCAACCCA ACCCAACCCA ACCCACCCA GTGCAGCCAA CTGGCAAATA  
 CGGTPTGGGT TGGTPTGGGT TGGTPTGGGT CACGTCGGTT GACCGTTTAT

4701 GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC  
 CAGAGGTGGG GGCCGTGATA GTGGCACTCA ACAGGCGTGG TGGCGTGCAG

4751 TCGCAGCCAA AAAAAAAAAA AGAAAGAAAA AAAAGAAAA GAAAAACAGC  
 AGCGTCGGTT TTTTTTTTTT TCTTTCTTTT TTTTCTTTTT CTTTTTGTG

4801 AGGTGGGTCC GGGTCGTGGG GGCCGGAAAA GCGAGGAGGA TCGCGAGCAG  
 TCCACCCAGG CCCAGCACCC CCGGCCTTTT CGCTCCTCCT AGCGCTCGTC

4851 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC  
 GCTGCTCCGG GCCGGGAGGG AGGCGAAGGT TTCTTTGCGG GGGGTAGCGG

4901 ACTATATACA TACCCCCCC TCTCCTCCA TCCCCCAAC CCTACCACCA  
 TGATATATGT ATGGGGGGGG AGAGGAGGGT AGGGGGGTTG GGATGGTGGT

4951 CCACCACCAC CACCTCCTCC CCCCTCGCTG CCGGACGACG AGCTCCTCCC  
 GGTGGTGGTG GTGGAGGAGG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG

5001 CCCTCCCCCT CCGCCGCCGC CGGTAACCAC CCCGCCCTC TCCTCTTCT  
 GGGAGGGGGA GCGCGCGGCG GCCATTGGTG GGGCGGGGAG AGGAGAAAGA

**FIG. 40G**



5051 TTCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT  
 AAGAGGCAAA AAAAAAAGCA GAGCCAGAGC TAGAAACCGG AACCATCAA

5101 GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG  
 CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

BamHI  
~~~~~

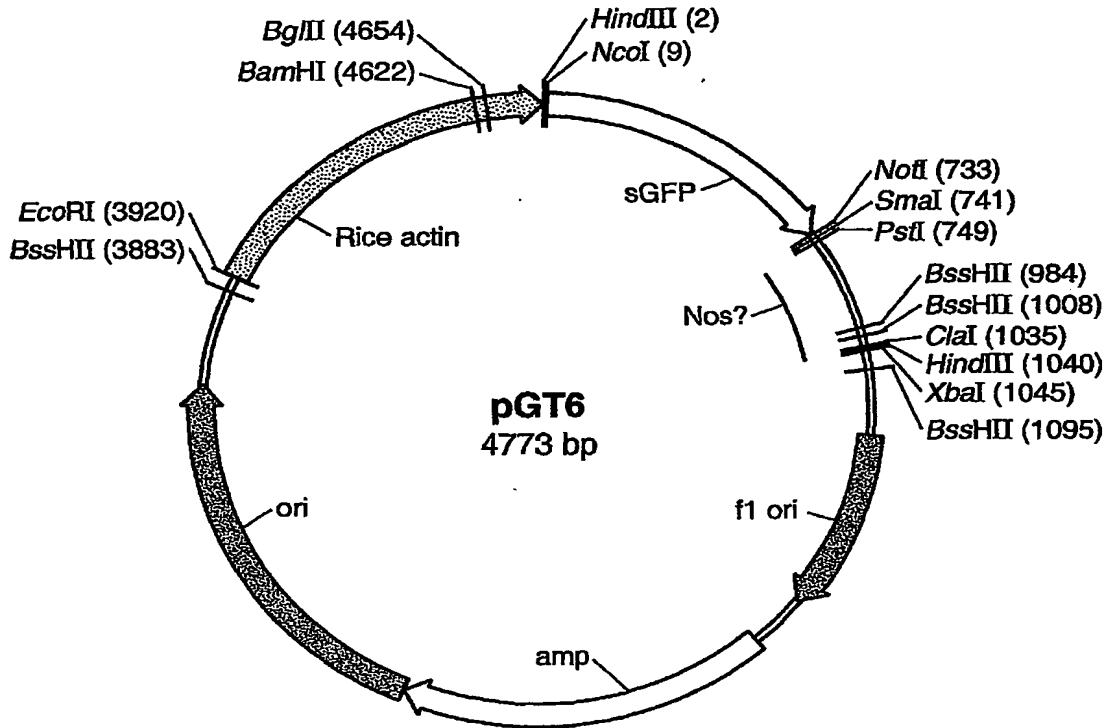
5151 GATCTCGCGG CTGGCGTCTC CGGGCCTGAG TCGGCCCGGA TCCTCGCGGG  
 CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG  
 CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AACACCATC

5251 AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT  
 TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC  
 CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

**FIG.\_40H**



**FIG. 41A**

HindIII NcoI  
 \*\*\*\*\*  
 1 AAGCTTACCA TGGTAGCAA GGGCGAGGAG CTGTTACCCG GGTGGTGGCC CATTCTGGTC GAGCTGGACC  
 TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAAGTGGC CCCACACAGG GTAGGACCAG CTCGACCTGC  
 71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGGGAT GCCACCTACG GCAAGCTGAC  
 CGCTGCACAT GCCGCTGTTT AAGTCGCACA GGGCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG  
 141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCAACC TCGTGACCAC CTTCACCTAC  
 GGACTTCAAG TAGACGTGGT GCGCGTTCGA CCGGCACGGG ACCGGTGGG AGCACTGGTG GAAGTGGATG  
 211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGCACTT CTTCAAGTCC GCCATGCCCG  
 CCGCACGTCA CGAAGTCCGC GATGGGCTG GTGTACTTCG TCGTGTGAA GAAATTCAGG CCGTACGGGC  
 281 AAGGCTACGT CCAGGAGCG ACCATCTTCT TCAAGGACGA CCGCAACTAC AAGACCCTGG CCGAGGTGAA  
 TTCCGATGCA GGTCCCTCGG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGGCG GGTCCACTT  
 351 GTTCGAGGGC GACACCCCTGG TGAACCCGAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CCGCAACATC  
 CAAGCTCCCG CTGTGGGACC ACTTGGGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG  
 421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCAGACAAG CAGAAGAACC  
 GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGTGTTTC GTCTTCTTGC  
 491 GCATCAAGGT GAACCTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTCGCCG ACCACTACCA  
 CGTAGTTCCA CTTGAAGTTC TAGCCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGTGGT  
 561 GCAGAACACC CCCATCGGC ACGGCCCGGT GCTGTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC  
 CGTCTTGTGG GGGTAGCCCG TGCCGGGCA CGACGACGGG CTGTGTGTTGA TGGACTCGTG GGTCAAGGCGG  
 631 CTGAGCAAAG ACCCCAACGA GAAGCGGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA  
 GACTCGTTTC TGGGGTTGCT CTTCCGCGTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG.-41B

```

SmaI
*****
NotI          PstI
*****
701  CTCACGGCAT GGACGAGCTG TACAAGTAAA GCGGCCGCC CCCTGCGAGG GAAACCACCTG AAGGATGAGC
    GAGTGCCCGTA CCTGCTCGAC AAGTTCATTT CGCCGGCGGG CCCGACGTC CTTTGGTGC TTTCTACTCG
771  TGTAAAGAAG CAGATCGTTC AAACATTTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
    ACATTTCTTC GTCAGCAAG TTTGTAACC GTTATTTCAA AGAATTTCTAA CTTAGGACAA CGGCCAGAAC
841  CGATGATTAT CATATAAATTT CTGTTGAAT ACGTTRAGCA TGTARAATTT AACATGTAAT GCATGACGTT
    GCTACTAATA GTATATTAAA GACAACTTAA TGCAATTCGT ACAATTATTA TTGTACATTA CGTACTGCCA
911  ATTTATGAGA TGGGTTTTTA TGATFAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA AAACAAAATA
    TAAATACTCT ACCCAAAAAT ACTAATCTCA GGGCGTTAAT AGTAAATTA TCGGCTATCT TTTGTTTTAT
          BssHII          ClaI HindIII
          *****          *****
981  TAGCGGCAA ACTAGGATAA ATTATCGGC GGGTGTCTAT CTATGTTACT AGATCGATAA GCTTCTAGAG
    ATCGCGCGTT TGATCCCTATT TAATAGCGG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC
          BssHII
          *****
1051 CGGCCGGTGG AGCTCCAATTT CGCCCTATAG TGAGTCGTAT TACGGCGCGT CACTGGCCGT CGTTTACAA
    GCCGGCCACC TCGAGGTTAA GCGGATATC ACTCAGCATA ATGGCGCGCA GTGACCGGCA GCAAAATGTT
1121 CGTCGTGACT GGGAAAACC TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCTT TTCGCCAGCT
    GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CCGAACGTCG TGTAGGGGGA AAGCGGTCCA
1191 GCGGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA
    CCGCATATATC GCTTCTCCGG GCGTGGCTAG CCGGAAGGTT TGTCAACCGG TCGGACTTAC CGCTTACCCCT
1261 CGGCCCTGTF AGCGGGCAT TAAGCGGCG GGGTGTGGTG GTTACGCGCA GCGTGAACCGC TACACTTGGC
    GCGCGGGACA TCGCCCGGTA ATTCGCGCG CCCACACCAC CAATGCGCGT CCGACTGGCG ATGTGAACGG

```

XbaI

\*\*\*\*\*  
ClaI HindIII  
\*\*\*\*\*

BssHII  
\*\*\*\*\*

BssHII  
\*\*\*\*\*

FIG. 41C

1331 AGGCCCTAG CGCCCGCTCC TTTGGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCGCCGTC TTTCGCCGTC  
 TCGCGGGATC GCGGGCGAGG AARGCGAAG AAGGGRAGGA AAGGCGGTG CAAGCGGCCG AAAGGGGCGAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT FCCGATTTAG TCGTTTACGG CACCTCGACC CCAAAAAACT  
 TTCCGAGATT AGCCCCCGAG GGAAATCCCA AGCCTAAATC ACGAAATGCC GTGGAGCTGG GGTTTTTTGA

1471 TGATTAGGGT GATGGTTTAC GTAGTGGGCC ATCGCCCTGA TAGACGGATT TTCGCCCTTT GACGGTTGGAG  
 ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT AFTGSCAAA AAGCGGAAA CTGCCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCTTATCTCG GTCTATTTCTT  
 AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATRAGAA

1611 TTGATTTATA AGGGATTTTG CCGATTTCCG CCTATTGGTT AAAAAATGAG CTGATTTTAA AAAAAATTAA  
 AACTAAATAT TCCCTAAAAC GSCATAAGCC GGATAACCAA TTTTTTACTC GACTAAATTTG TTTTTTAATTT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTTAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC  
 CGCCTTAAA TTGTTTTTATA AFTCGAATG TTAAATCCAC CGTGAAAAAG CCCTTTACAC CGCCTTTGGG

1751 CTATTGTGTT ATTTTTCTAA ATACATCAA ATATGTATCC GCTCATGAGA CAATAACCTT GATAAATGCT  
 GATAAACAAA TAAAAGATT TATGTAAAGT TATACATAGG CGAGTACTCT GTTATTTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCCGTGTG CCCTTATTTCC CTTTTTTTGG  
 AGTTATTTATA ACTTTTTTCTT TCTCATACTC ATAAAGTTGTA AAGGCCACAGC GGGAAATGAG GAAAAAACGC

1891 GCATTTTGGC TTCCCTGTTTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG  
 CGTAAAACGG AAGGACAAA ACGAGTGGGT CTTTTGGACC ACTTTTCAATT TCTACGACTT CTAGTCAACC

1961 GTGCACGAGT GGGTTACATC GAACCTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTC GCCCCGAAGA  
 CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCCC AFTCTAGGAA CTCTCAAAAAG CCGGGCTTCT

2031 ACGTTTTCCA ATGATGAGCA CTTTTAAAGT TCTGTATGTT GCGCGGGTAT TATCCGTTAT TGACGCCGGG  
 TGCAAAAGGT TACTACTCGT GAAAATTTCA AGACGATACA CCGCGCCATA ATAGGCAATA ACTGCGGCC

2101 CAAGAGCAAC TCGGTGCGCG CATACACTAT TCTCAGAAAG ACTTGGTTGA GTACTACCCA GTCACAGAAA  
 GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

FIG. 41D

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCCTAC TGTCATCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTCTGACAA CGATCGAGG ACCGAAGGAG CTAACCCGCTT TTTGACAAA CATGGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCCTCC TGGCTTCCCTC GATTGGCGAA AAAACCTGTT GTACCCCTTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
GTACATTTGAG CCGAACTAGC AACCCCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACCTGTGGT

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGSGGAA CTACTTACTC TAGCTTCCC  
GCTACGGACA TCGTTACCCTG TGTTCACACG CGTTTGATAA TTGACCCGCTT GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGGGGA TAAAGTTGCA GGACCACCTTC TGGCTCCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCCCT ATTTCACCGT CCTGGTGAAG ACGCGAGCCG GGNAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTTGCA GCACITGGGGC  
CCGACCAAAAT AACGACTATT TAGACCTCCG CCACTCCGAC CCAGAGCCGC ATAGTAAAGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCTGATTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA  
GTCTACCATT CGGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGTCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCCTCACT GATTAAGCAT TGGTAACTGT CAGACCAGT TTACTCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTAAA ACTTCATTTT TAAATTAATA GGATCTAGGT GAAGATCCTT TTTGATTAATC  
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAAATTTT CCTAGATCCA CTTCTAGGAA AACTATATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CPTCCACTG AGCTCAGAC CCCGTAGAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAAAT GCACCTCAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCCTTTT TTCTGGCGGT AATCTGCTGC TTGCAACAA AAAAACCCACC GCTACCAGCG  
TAGAAGAAT CTAGGAAAAA AAGACCGGCA TTAGACGACG AACGTTTGT TTTTTTGGTG CGATGGTCCG

2941 GTGGTTTGT TCCCGGATCA AGAGTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
CACCAACAA ACGGCCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAATCG TCTCGCGTCT

FIG.- 41E

3011 TACCAAATAC TGTCCCTTCTA GTGTAGCGGT AGTTAGGCCA CCACCTCAAG AACTCTGTAG CACCGCCTAC  
 ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGTCTCTGCC AGTGGCGATA AGTCTGTCT TACCGGGTTG  
 TATGGAGCGA GACGATTAGG ACAATGTCA CCGACGACGG TCACCCGTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTCCG GCTGAACGGG GGTTCGTGC ACACAGCCCA  
 CTGAGTCTG CTATCAATGG CCTATTCGC GTCCCCAGCC CGACTTGCCT CCCAAGCACG TGTGTCCGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
 CGAACCTCCG TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GTTCCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCCAC GAGGGAGCTT  
 GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGT CTTCTCGGTG CTCCCTCGAA

3361 CCAGGGGAA ACGCCTGGTA TCTTTAAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CFTCGATTTT  
 GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAATC GCAGCTAAAA

3431 TGTGATGCTC GTCAGGGGG CGGAGCTAT GGAATAACGC CAGCAACGG GCTTTTATC GGTTCCTGGC  
 AACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCTGTTCGC CGGAAAAATG CCAAGGACCG

3501 CTTTGTCTGG CCTTTTGTCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTCTGGATAA CCGTATTACC  
 GAAAAACGACC GGAATAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGTCTGCCG AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG  
 CGGAAACTCA CTCGACTATG GCGAGCGCG TCGGCTTGT GGTCTGCCGC GCTCAGTCAC TCGCTCCTTC

3641 CGGAAGAGCG CCCAATACGC AACCCGCTC TCCCCTGCGG TTGGCCGATT CAFTAATGCA GCTGGCACGA  
 GCCTCTCCG GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGGCTAA GTAATTACGT CGACCCGTGCT

3711 CAGGTTTCCC GACTGGAAAG CCGGCAAGTA GCGCAACGCA ATTAATGTA GTTAGCTCAC TCATTAGGCA  
 GTCCAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAAATGT GAGCGGATAA CAATTCACA  
 GGGTCCGAA ATGTGAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCTATT GTTAAAGTGT

FIG.-41F

```

      B55HII                                     EcorI
      ~~~~~
3851  CAGGAACAG CTATGACCAT GATTAGCCA GATTAGCCA AGCGGCAAT TAACCTCAC TAAAGGAAAC AAAAGCTGGA
      GTCCTTTGTC GATACTGGTA CTAATGCCGT TCGCGCGTTA ATYGGAGTG ATTTCCCTTG TTTTCGACCT
      EcorI
      ~~~~~
3921  ATTCCACAAT GAACAATAAT AAGATTAATA TAGTTGCC CCCTGCAGC GATGGTAAT TTTTCTAGTA
      TAAGTGTTA CTTGTTATTA TTCTAATTTT ATCGAACGGG GGCAACGTCG CTACCCATAA AAAAGATCAT
      ~~~~~
3991  AATAAARA AACTTTAGA CTCAAAACAT TTACAAAAC AACCCCTAAA GTCTAAAGC CCAAAAGTCT
      TTTATTTTCT ATTTGAATCT GAGTTTGTG AATGTTTTG TTGGGGATTT CAGGATTCG GGTTCACGA
      ~~~~~
4061  ATGCACGATC CATAGCAAGC CCAGCCCAAC CCAACCCAAC CCAACCCACC CCAGTGCAGC CAACTGGCAA
      TACGTGCTAG GTATCGTTTC GGTCCGGTTG GGTTCGGTGG GGTTCAGTCC GGTTCAGTCCG GTTGACCGTT
      ~~~~~
4131  ATAGTCTCCA CCCCAGCAC TATCACCGTG AGTTGTCCGC ACCACCGCAC GTCTGCAGC CAAAAAATA
      TATCAGAGGT GGGGCGCGTG ATAGTGGCAC TCAACAGCGG TGGTGGCGTG CAGAGCGTCG GTTTTTTTTT
      ~~~~~
4201  AAAAAGAAGA AAAAAAGAA AAAGAAAAC AGCAGGTGGG TCCGGTCTGT GGGGGCCGGA AAAGCGAGGA
      TTTTCTTTCT TTTTCTTTCT TTTTCTTTTG TCGTCCACCC AGCCCCAGCA CCCCCGGCCT TTTTCGTCCT
      ~~~~~
4271  GGATCGCGAG CAGCGACGAG GCCCGGCCCT CCTCCGCTT CCAAGAAAC GCCCCCCATC GCCACTATAT
      CCTAGCGCTC GTCGCTGCTC CGGGCCGGGA GGGAGCGGAA GGTTCCTTTG CGGGGGTAG CGGTGATATA
      ~~~~~
4341  ACATACCCCC CCTCTCCTC CCAATCCCCC AACCTTACCA CCACCACCAC CACCACCTCC TCCCCCTCG
      TGTATGGGGG GGGAGAGGAG GGTAGGGGGG TTGGGATGGT GGTGGTGGTG GTGTGGAGG AGGGGGAGC
      ~~~~~
4411  CTGCCGGACG ACGAGTCTCT CCCCCTCCC CCTCCGCTCG CGCCGGTAAC CACCCGGCC CTCTCTCTTT
      GACGGCCTGC TGCTCGAGGA GGGGGGAGGG GAGGGCGCG GCGGCCATTG GTGGGGCGGG GAGAGGAGAA
      ~~~~~
4481  TCTTCTCCG TTTTTTTTTT CGTCTCGTC TCGATCTTTG GCCTTGGTAG TTTGGGTGG CGAGAGCGGC
      AGAAAGAGGC AAAAAAARA GCAGAGCCAG AGCTAGAARAC CGGAACCATC AAACCCACC GCTCTCGCCG
      ~~~~~
4551  TTCGTGCCCC AGATCGGTGC GCGGGAGGGG CGGGATCTCG CCGTGGCGT CTCGGGGCGT GAGTCGGCCC
      AAGCAGCGGG TCTAGCCACG CGCCCTCCC GCCCTAGAGC CCGGACCGCA GAGGCCGCA CTCAGCCGGG

```

FIG.- 41G

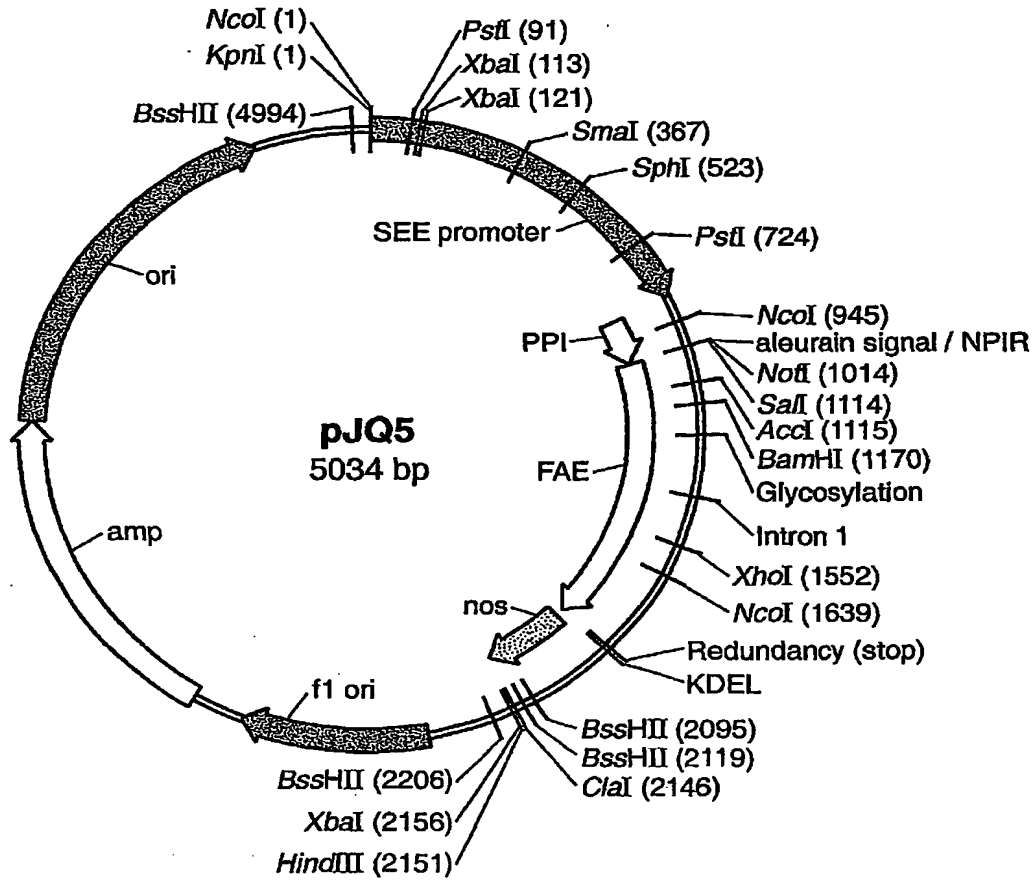


BamHI  
 ~~~~~  
 4621 GGATCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTTCTTTCTT CTTTTTGGG TAGAATTTGA  
 CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAAG AAAGAAAGAA GAAAAACACC ATCTTAAACT

BglII  
 ~~~~~  
 4691 ATCCCACAGC ATTGTCATC GGTAGTTTTT CTTTTCATGA TTTGTGACAA ATGCAGCCTC GTCCGGAGCT  
 TAGGGAGTCG TAAACAAGTAG CCATCAAAA GAAAGTACT AAACACTGTT TACGTCGGAG CACGCCCTCGA

4761 TTTTGTAGG TAG  
 AAAACATCC ATC

FIG. 41H



**FIG. 42A**

```

NcoI
-----
KpnI
-----
1  CATGGCCAG  GTATAATTAT  GGGATATCTC  AAGCAATAAA  TCGAAATATC  ACCATGGGCT  ACAATATCTG
   GTACCCGGTC  CATAATTAATA  CCCTATAGAG  TTCGTTTATT  AGCTTTATAG  TGGTAACCGA  TGTATATAGAC

                               PstI                               XbaI                               XbaI
                               -----                               -----                               -----
71  AGCTCCGAGT  TCTGACTGCA  GTCTGGATGA  CGGRTGTGT  ATCTAGAACT  CTAGATAGCA  CAGCCACAGC
   TCGAGGCTCA  AGACTGACGT  CAGACCTACT  GCGCACAAAC  TAGATCTTGA  GATCTATCGT  GTCGGTGTCC

141  ACCTACAGGA  GTGCGACACT  TGTGGACTGT  AGTAGTGTTG  GAGACGGAGC  TCTTTCCTAC  CTCTGACGTT
   TGGATGTCTT  CACGCTGTGA  ACACCTGACA  TCATCACAAAC  CTCTGCCTCG  AGAAAGGATG  GAGGACTGCA

211  TGCCGCCGTT  GTCCATTCCA  ACGGCATCAC  TCTCAACCAA  TCACGGGCTC  CCAACAAAAT  ATCGTCCCCC
   ACGGCGGCAA  CAGGTAAGGT  TGCCGTAGTG  AGAGTTGGTT  AGTGCCGGAG  GGTGTGTTTTA  TAGCAGGGGG

281  ATGTCTTGGC  GGAGAGAGAG  TACATACATG  CTGTCCGGCC  GTTTTTGTCT  GAATCTCGCT  FCCACTGGCC
   TACAGAAACCG  CCTCTCTCTC  ATGTATGTAC  GACAGGCGGG  CAAAACACAG  CTTAGAGCGA  AGGTGACC GG

                               SmaI
                               -----
351  AATCAGCTCA  GCTCCCGGGA  GCTCACTCAT  TCAAGATCCC  ATCGTCGTCC  TCACCCCTGG  CGTCATGGGA
   TTAGTCGAGT  CGAGGGCCCT  CGAGTGAGTA  AGTTCTAGGG  TAGCAGCAGC  AGTGGGGACC  GCAGTACCCCT

421  TGGAAAAGAA  CCTCCGTTGC  TCGGATGAGT  CAGCCATATC  CCCGAACAGA  GTACTGCAAG  ATAACCCCAAT
   ACCTTTCTTT  GGAGGCAACG  AGCCTACTCA  GTCCGTATAG  GGGCTTGTCT  CATGACGTTT  TATTTGGGTTA

                               SphI
                               -----
491  TCAGATTCCC  CCAATAGAGA  AAGTATAGCA  TGCTTCCGGG  TTTTGTTTGG  CTTAATGAC  TTTATTTTGT
   AGTCTAAGGG  GGTATCTCT  TTCATATCGT  ACGAAAGCCC  AAACAAACC  GAATTAACG  AAATAAAAC

561  TTGGAGTTGA  ATGCTGATTT  GTTGTGTAAG  ATGCCAACCC  ATCTGAATAT  CGAGACGGAT  AATAGGCTGG
   AACCTCAACT  TAGACTAAA  CAACACATTT  TACGGGTTGG  TAGACTTATA  GCTCTGCCTA  TTATCCGACC

```

FIG. 42B

631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCAATTACA GCTGGAGGCT  
 GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI

\*\*\*\*\*

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAALAC  
 AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTTACCC ACTTCGCACC GCTACTCTAC CCAATATTTG

771 CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCACTACCAT CTCGCCCTGC TCCCCCTGCC GGACGACCCA  
 GGGCCGTTGG CCTTGCCTC GAGGGCGGAT GGTCAATGGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAATAACT GTTGCCCACT CGCCGGCGAG ATGCMCGTGC ACAAGGAGGT SAACCTCTGTS GCCTACCTCC  
 CAFTTATGA CAACGGGTGA GCGGCCGCTC TACCKGCACG TGTTCCTCCA STTGAAGCAS CGGATGGAGG

NcoI

\*\*\*\*\*

911 TGATCGTSC TCGGCTCCTC TTGCTCGTST CCGCCATGGA GCACGTGGAC GCCAAGGCTT GCACCCCKGA  
 ACTAGCABGA GCCGGAGGAG AACGAGCASA GCGGTACCT CGTGCACCTG CGFTCCCGA CGTGGGCGCT

NotI

\*\*\*\*\*

981 GTCCGGCAAC CTCGGCTTCG GCATCTGCC GCGGCGCCGC TCCACGCAGG GCATCTCGA AGACCTCTAC  
 CACGCCGTTG GAGCCGAAGC CGTAGACGGG CCGCCGCGGG AGGTCCCTCC CFTAGAGGCT TCTGGAGATG

SalI

\*\*\*\*\*

AclI

\*\*\*\*\*

1051 AGCCGTTTAG TCGAAATGGC CACTATCTCC CAAGCTGCCT ACGCCGACCT GTGCAACATTT CCGTCCGACTA  
 TCGGCAAAATC AGCTTTACCG GTGATAGAGG GTTCGACGGA TCGCGGCTGA CACGTTOTAA GGCAGCTGAT

BamHI

\*\*\*\*\*

1121 TTATCAAGGG AGAGAAAATT TACAATCTC AAATGACAT TAACGGATGG ATCCTCCGGG ACGACAGCAG  
 AATAGTTCCC TCTCTTTTAA ATGTTAAGAG TTTGACTGTA ATTGCCTACC TAGGAGGGCG TGCTGTCTGC

FIG. 42C

1191 CAAGAAATA ATCACCGTCT FCCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACATACACC  
 GTTTCATTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TECTTAGATG TTGAGCTATG AATTGATGTGG  
  
 1261 CTCACGCCCT TCGACACCCCT ACCACAATGC AACGTTGTG AAGTACACCG TGGATATTAT ATTGGATGGG  
 GAGTGGGGA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTCATGTGCC ACCTATAATA TAACCTACCC  
  
 1331 TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTAGCCAG TATCCGGACT ACGCGCTGAC  
 AGAGGCAGGT CCTGGTTCAG CTCAGCGAAC AGTTTTCGT CCAATCGGTC ATAGGCCCTGA TCGCGGACTG  
  
 1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC  
 GCACGTGGCCG GTGMMGGAGC CCGGGAGGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTATGCTG  
  
 1471 AACATCCGCC TGTACACCTT CGGCGAACC GCGAGCGCA ATCAGGCCCTT CCGGTCTGAC ATGAACGATG  
 TTGTAGGCCG ACATGTGGA GCCGCTTGGC GCCTCGCGT TAGTCCGGAA GCGCAGCATG TACTTGTCTAC  
  
 XhoI  
 \*\*\*\*\*  
 1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCCG GGTCACTCAT GCCAACGACG GCATCCCAAA  
 GGAAGTTCG GAGCTCGGGT CTATGCTCGG TCATAAAGC CCAATGAGTA CCGTTGCTGC CGTAGGGTTT  
  
 NcoI  
 \*\*\*\*\*  
 1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC  
 GBACGGGGC CACCTCGTCC CCATGCGGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTGCGCG  
  
 1681 CAGAACACAT TTGTCTGCAC TGGGGATGAA GTGCAGTGT GTGAGGCCCA GGGCGGACAG GGTGTGAATA  
 GTCTTGTGTA AACAGACGTG ACCCCTACTT CACGTACAGA CACTCCGGGT CCGCCCTGTC CCACACTTAT  
  
 1751 ATGGCCACAC GACTTATTTT GGGATGACGA GCGGACCTG TACATGTTGA TCAGTCAFTT CAGCCTCCCC  
 TACGCGTGTG CTGAATAAAA CCTTACTGCT CGCCTCGGAC ATGTACCCT AGTCAGTAAA GTCGGAGGGG  
  
 1821 GAGTGTACCA GGAAGATGG ATGTCTTGA GAGGGGCCG CGTAACCACT GAAGGATGAG CTGTAAAGAA  
 CTCACATGGT CCTTTCWACC TACAGGACCT CTCGCCCGC GCATTTGGTGA CTTCCCTACTC GACATFTCTT

FIG.- 42D

1891 GCAGATCGTT CAAACATTTG GCAATAAAGT TTCTTAAAGT TGAATCCTGT TGCCGGTCTT GCGATGATTA  
 CGTCTAGCAA GTTTGTAAAC CGTTATTTCA AAGAAATCTA ACTTAGGACA ACGGCCAGAA CGCTACTAAT

1961 TCATAATAAT TCTGTTGAAT TACGTTAAGC ATGTAATAAT TAACATGTAA TGCATGACGT TATTTATGAG  
 AGTATATATA AGACAACCTTA ATGCAATTCG TACATATATA ATTGTACAT ACGTACTGCA ATAAATACTC

BssHII  
\*\*\*\*\*

2031 ATGGGTTTTT ATGATTAGAG FCCCGCAAT ATACATTTAA TACGGATAG AAAACAAAT ATAGCGCGCA  
 TACCCAAAA TACTAATCTC AGGGCGTTAA TATGTAAAT ATCGCTATC TTTTGTTTTA TATCGCGCGT

XbaI  
\*\*\*\*\*

BssHII  
\*\*\*\*\*  
ClaI HindIII  
\*\*\*\*\*

2101 AACTAGGATA AATTAATCGG CGCGGTGCA TCTAATGTTAC TAGATCGATA AGCTTCTAGA GCGCGCGGTG  
 TTGATCCTAT TTAATAGCGC GCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT CGCCGGCCAC

BssHII  
\*\*\*\*\*

2171 GAGCTCCAAAT TCGCCCTATA GTAGTCGTA TTACGGCGC TCACATGGCCG TCGTTTTACA ACGTCGTGAC  
 CTCGAGGTTA AGCGGATAT CACTCAGCAT AATGCGCCCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241 TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCA CACATCCCC TTTCCGCCAGC TGCCCTAATA  
 ACCCTTTTGG GACCGCAATG GGTGAATTA GCGGAACGTC GTGTAGGGGG AAAGCGGTG ACCGCATAT

2311 GCGAAGAGC CCGCACCGAT CGCCCTTCCC AACAGTTGG CAGCCTGAT GCGGAATGG ACGCGCCCTG  
 CGCTTCTCCG GGCCTGGCTA GCGGGAAGG TTGTCAACGC CTCGACTTA CCGCTTACC TCCCGCGGAC

2381 TAGCGGCGCA TTAAGCGCGG CGGGTGTGGT GGTACGCGC AGCGTGACC CTACACTGG CAGCGCCCTA  
 ATCGCCGCGT AATTCGGCC GCCCACACCA CCAATGCGCG TCGCACTGGC GATGTGAACG GTCGCGGGAT

2451 GCGCCCGCTC CTTTCGCTTT CTTCCTTCC TTTCTGCGCA CGTTCGCCGG CTTTCCCGGT CAAGCTCTAA  
 CGCGGGCGAG GAAGCGAAA GAAGGGAAGG AAAGAGCGGT GCAAGCGGCC GAAGGGGCA GTTCGAGATT

FIG.-42E

2521 ATCGGGGGCT CCTTTAGGG TTCCGANTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG  
 TAGCCCCCGA GGGAAATCCC AAGCTAAAT CACGAAATGC CGTGGAGCTG GGGTTTTTG AACTAATCCC

2591 TGATGGTTCA CGTAGTGGC CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTT  
 ACTACCAAGT GCATCACCCG GTAGCGGAC TATCTGCCAA AAGCGGGA ACTGCACCT CAGGTGCAAG

2661 TTTAATAGTG GACTCTTGT CCAACTGGG ACAACACTCA ACCCTATCT GGTCTATCT TTTGATTTAT  
 AAATATCAC CTGAGAACA GFTTGGACCT TGTGTGAGT TGGGATAGG CCAGATAAGA AAATAATA

2731 AAGGGATTTT GCCGATTCG GCCTATGGT TAAAAAAGA GCTGATTTA CAAAAATTA ACGGAAATTT  
 TTCCCTAAA CCGCTAAAAG CGGATAACCA ATTTTTTACT CGACTAAAT GTTTTTAAAT TGGCTTAAA

2801 TAACAAAATA TTAACGCTTA CAATTTAGT GGCCTTTTC GGGGAAATGT GCGCGAACC CCTATTGT  
 APTGTTTTAT AATTGCCAAT GTTAAATCCA CCGTGAAAAG CCCCTTTACA CGCGCTTG GATATAACAA

2871 TATTTTTCTA AATACATTCA AATATGTAT CGTCTATGAG ACAATAACC TGATAAATGC TTCAATAATA  
 ATAAAAAGAT TTATGTAAAT TTATACATAG GCGAGTACT TGTATTGGG ACTATTTACG AAGTTATTAT

2941 TTGAAAAAGG AAGAGTATGA GTATTCALCA TTTCGGTGT CCCCATTTC CCTTTTTGC GGCATTTGC  
 AACTTTTTCC TTCTCATACT CATAGTTGT AAAGCACAG CCGGAAATAG GAAAAAACC CCGTAAAACG

3011 CTTCCGTGTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
 GAAGGACAAA AACGAGTGGG TCTTTGCCAC CACTTTCAAT TTCTACGACT TCTAGTCAAC CCACGTGCTC

3081 TGGGTTACAT CGAACTGGAT CTCACACGCG GTAAGATCCT TGAGAGTTTT CGCCCCGAAG AACGTTTTCC  
 ACCCAATGTA GCTTGACCCTA GAGTTGTCCG CATTCTAGGA ACTCTCAAAA GCGGGCTTC TTGCAAAAGG

3151 AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA TTRTCCCGTA TTGACGCCGG GCAAGAGCAA  
 TTACTACTCG TGAAAAATTTC AAGACGATAC ACCGCGCCAT AATAGGGCAT AACTGCGGCC CGTTCCTCGTT

3221 CTCGGTCGCC GCATACACTA TTCTCAGAA T GACTTGGTTG AGTACTCACC AGTCACAGAA AAGCATCTTA  
 GAGCCAGCGG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGAAT

3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT GATAACACTG CGGCCAACTT  
 GCCTACCGTA CTGTCATTCT CTTAAATACGT CACGACGGTA TTGGTACTCA CTATTTGAC GCCGGTTGAA

FIG.-42F

3361 ACTTCGACA ACGATCGGAG GACCGAAGGA GCTAACCCGCT TTTTGGCACA ACATGGGGGA TCATGTAAC<sup>T</sup>  
 TGAAGACTGT TGCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAAACGTGT TGTACCCCTT AGTACATGGA  
 3431 CGCCTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCCTGCACCC ACGATGCCTG  
 GCGGAAC<sup>T</sup>AG CAACCCCTGG CCTCGACTTA CTTCGGTAG GTTTGCTGCT CGCACTGTGG TGCTACGGAC  
 3501 TAGCAATGGC AACAAAGTTG CGCAAAC<sup>T</sup>AT TAACTGGCGA ACTACTTACT CTAGCTTCCC GGCAACAAT<sup>T</sup>  
 ATCGTTACCG TTGTTGCAAC GCGTTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CCGTTGTTAA  
 3571 AATAGACTGG ATGGAGGCGG ATAAAGTTGC AGGACCAC<sup>TT</sup> CTGCGCTCGG CCTTCCGGC TGGCTGGTT<sup>T</sup>  
 TTATCTGACC TACCTCCGCC TATTTCAACG TCCTGGTGAA GACCGGAGCC GGGAAAGGCC ACCGACCAAA  
 3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGCG GTATCAT<sup>T</sup>GC AGCACTGGGG CCAGATGGTA  
 TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG TCGTGCACCC GGTCTACCAT  
 3711 AGCCCTCCC TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
 TCGGGAGGGC ATAGCATCAA TAGATGTGCT GCCCTCAGT CCGTTGATAC CTACTTGTCTT TATCTGTCTA  
 3781 CGCTGAGATA GGTGCCTCAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA TATACTTTAG  
 GCGACTCTAT CCACGGAGTG ACTAATTCGT AACCATGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC  
 3851 ATTGAATTAA AACTTCATTT TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGAATAAT CTCATGACCA  
 TAACTAAAAT TTGAAGTAAA AATTAATTT TCCTAGATCC ACTTCTAGGA AAAACTAATA GAGTACTGGT  
 3921 AAATCCCTTA ACGTGAGTTT TCGTTC<sup>CACT</sup> GAGGTCAGA CCCCCTAGAA AAGATCAAAG GATCTTCTTG  
 TTTAGGGAA<sup>T</sup> TGCAC<sup>TCAA</sup> AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTC<sup>TAGTTTC</sup> CTAGAAGAAC  
 3991 AGATCC<sup>TTTT</sup> TTTCTGCGCG TAATCTGCTG CTTCGAACA AAAAACCAC CGCTACCAGC GGTGGTTGT  
 TCTAGGAAA AAAGACCGC ATTAGACGAC GAACGTTGT TTTTTTGGTG GCGATGGTCC CCACCAACA  
 4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGNAGSTAA CTGGCTTCAG CAGAGCGCAG ATACCAATA  
 AACGGCCTAG TTCTCGATGG TTGAGAAAAA GCGTCCAT<sup>T</sup> GACCGAAGTC GTCTCCGCTC TATGGTTTAT  
 4131 CTGTCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTCAA GAAC<sup>TCTGA</sup> GCACCGCCTA CATACCTCGC  
 GACAGGARGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

FIG. 42G



4201 TCTGCTAATC CTGTTTACCAG TGGCTGCTGC CAGTGGCCAT AAGTCGTGTC TTACCGGGTT TACTCAAGA  
AGACGATTAG GACAATGGTC ACCGACGACG GTACCCGCTA TTCAGCACAG AATGGCCCCAA CCTGAGTTCT

4271 CGATAGTTAC CGGATAAGGC GCAGCGGTCG GCTGTAACGG GGGGTTGCTG CACACAGCCC AGCTTGGAGC  
GCTATCAATG GCCTATTCGG CGTGCCCAGC CCGACTTCCC CCCC AAGCAC GTGTGTGCGG TCGAACTTCG

4341 GAACGACCTA CACCGAACTG AGATACCTAC AGCTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG  
CTTGTGAGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTTCG CGGTGCGAAG GGCCTCCCTC

4411 AAAGGGGAC AGGTATCCGG TAAGCGGCAG GGTCCGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGGG  
TTTCCGGCTG TCCATAGGCC ATTCCGCCGTC CCAGCCTTGT CCTCTCCGT GCTCCCTCGA AGGTCCCCCT

4481 AACGCTGGT ATCTTTATAG TCTGTGCGG TTTCCGACC TCTGACTTGA GCGTCGATTT TTGTGATGCT  
TTGCCGACCA TAGAATATC AGGACAGCCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA

4551 CGTCAGGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCCCTGG CCTTTTGTG  
GCAGTCCCCC CGCCTCGGAT ACCTTTTTCG GGTGCTTCCG CCGGAAAAAT GCCAAGGACC GGAARAACGAC

4621 GCCTTTGTCT CACATGTTCT TTCTTGGCTT ATCCCTTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG  
CGGAAACGA GTGTACAAGA AAGGACGCCA TAGGGACTA AGACACCTAT TGGCATAATG GCGGAAACTC

4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGA GCGGAAGAGC  
ACTCGACTAT GCGGAGCGGC CTCGGCTTGC TGGCTCGGCT CGCTCAGTCA CTCGCTCCTT CGCCTTCTCG

4761 GCCCAATACG CAAACCGCCT CTCCTCCGCG GTTGGCCGAT TCATTAATGC AGCTGGCAGC ACAGGTTTCC  
CGGTTATGC GTTTGGCCGA GAGGGCCGCG CAACCGGCTA AGTAAATTACG TCGACCCTGC TGTCCAAAGG

4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
GCTGACCTTT CGCCCCGTCAC TCGCGTTTGG TTAATTAAC TCAATCGAGT GAGTAACTCG TGGGTCCCGA

4901 TTACTCTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATG TGAGCGGATA ACAATTTTAC ACAGGAAACA  
AATGTGAAT ACGAAGGCCG AGCATAACAC ACACCTTAAC ACTCCCTAT TGTATAAGTG TGTCTTTTGT

FIG.- 42H

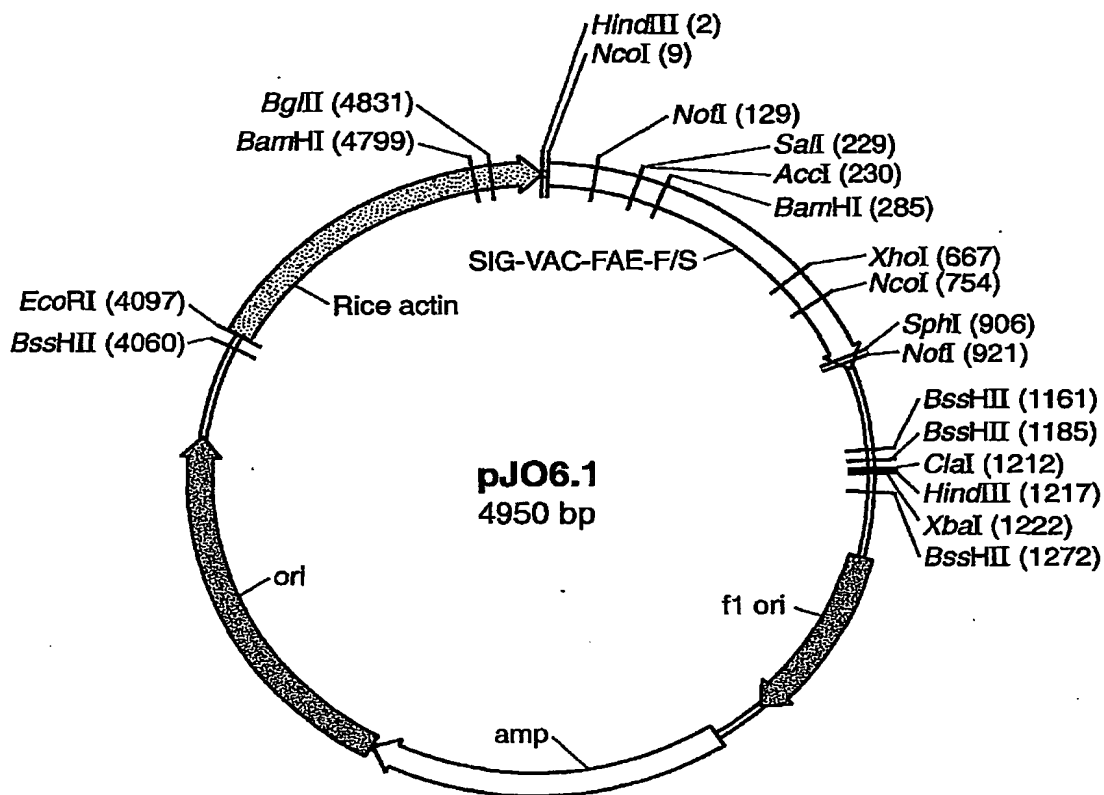
4971 GCTATGACCA TGATTACGCC AAGCGCGCA TTAACCTCA CTAAGGGA CAAAGCTGG GTAC  
CGAFACTGGT ACTAATGCGG TTCGCGCGTT AATTGGAGT GATTCCTT GTTTCGACC CATG

BssHI  
\*\*\*\*\*

KpnI  
\*\*\*\*\*

NcoI

**FIG. 421**



**FIG. 43A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGCCACGC CCGGTCTTC CTCCTGGCG TCGCCGTGCT GGCCACGGCC GCCGTGGCC  
 TTCGAATGGT ACCGGTGG GCGCAGGAG GAGGACCCG AGCCGCACGA CCGGTGCCG CGGCACGGC

NotI  
 ~~~~~  
 71 TCGCCTCCTC CTCCTCTTC GCCGACTCCA ACCGATCCG GCCGTGACC GACCGCGCG GCGCCTCCAC  
 AGCGGAGGAG GAGGAGGAA CCGCTAGGT TGGCTAGGC CCGCAGTGG CTGGCCGCC GCGGAGGATG

141 GCAGGGCATC TCCGAGACC TCTACAGCC TTTAGTCGA ATGCCACTA TCTCCCAAG TGCCTACGCC  
 CGTCCCCTAG AGGCTCTGG AGATGTCGC AATCAGCTT TACCGGTGAT AGAGGGTTC ACGGATGCGG

Sall  
 ~~~~~  
 211 GACCTGTGCA ACATTCCTC GACTATTATC AAGGAGAGA AAATTTACA TTCTCAAAT GACATTAACG  
 CTGGACACGT TGTAAAGCAG CTGATAATAG TTCCCTCTCT TTTAAATGTT AAGAGTTGA CTGTAAATTC

BamHI  
 ~~~~~  
 281 GATGGATCCT CCGCGAGGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCAC TGGA GTGATACGAA  
 CTACCCTAGGA GCGCTGCTG TCGTCTGTTT TTTATTAGTG GCAGAAAGCA CCGTGACCAT CACTATGCTT

351 TCTACAATC GATCTAAT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA  
 AGATGTTGAG CTATGATTGA TGTGGGATG CGAAAGCTG TGGATGGTG TTACGTTGCC AACACTTCAT

421 CACGGTGGAT ATTATATTGG ATGGTCTCC GTCCAGGACC AGTCCAGTC GCTTGTCAA CAGCAGGTTA  
 GTGCCACCTA TAATATAAC TACCAGAGG CAGGTCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT

491 GCCAGTATCC GGACTACGG CTGACCCTGA CCGGCCACK CCTCGCGCC TCCCTGGCG CACTCACTGC  
 CCGTCAATAG CCTGATGCG GACTGGCACT GCGCGGTGAG GAGCCCGGG AGGACCGCC GTGAGTGACG

561 CGCCAGCTG TCTGGACAT ACGACAACAT CCGCCTGTAC ACCCTCGCG AACCGCGCAG CGGCAATCAG  
 CCGGCTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG TGAAGCCG TTGGCGCGTC GCCGTTAGTC

FIG. 43B

XhoI

631 GCCTTCGGGT CGTACATGAA CGATGCCCTTC CAAGCCCTGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
 CGGAAGCCCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CCGGTCTATG CTGGTCATA AAGGCCCAGT

NotI

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCTAG CCGGTACC GC CATCTCAT  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGC TGCAC TGGG ATGAA GTGCA GTGCTGTGAG  
 GACCTCGCA CTAGGAATGT CGCGGTCTT GTGTAAACAG ACGTGACCC TACTTCACGT CACGACACTC

SphI

841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT APTTTGGGAT GACGAGCGGC GCATGCACCT  
 CCGGTCCCBC CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCCTA CTGCTCGCCG CGTACGTGGA

NotI

911 GGCCGGTCGC GCGCGCGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CAFTTGGCAA  
 CCGGCCAGCG CCGCGCCCTT TGGTGACTTC CTACTCGACA TTCTTTCGTC TAGCAAGTTT GTAAACCCTT  
 981 TAAAGTTCT TAAAGTTGAA TCCTGTTGCC GGTCTTGGG TGAATTATCAT ATAATTTCTG TTGAATTTAG  
 ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTARAGAC AACTTAATGC  
 1051 TTAAGCATGT AATAATTAAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTTATGA TTAGAGTCCC  
 AATTTCGTACA TTATTAATTTG TACATTACGT ACTGCAATAA ATACTCTACC CAAAATACT AATCTCAGGG

B88HII B88HII  
 \*\*\*\*\*

1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAAATATAG CCGCACAAC AGGATAAAT ATCGCGCGCG  
 CGTTAATATG TAAATTATGC GCTATCTTTT GTTTTATATC GCGGTTTGA TCCTATTTAA TAGCGCGCGC

FIG. 43C

XbaI

Clal HindIII

1191 GTGTCATCTA TGTACTAGA TCGATPAGCT TCTAGAGCGG CCGGTGGAGC TCCAAFTCGC CCTATAGTGA  
CACAGTAGAT ACAATGATCT AGCTATTCTGA AGAFTCTGCC GSCCACCCTCG AGGTTAAGCG GGATATCACT

BssHII

1261 GTCGTATTAC GCGCGTCCAC TGGCCGTCTGT TTTACAACGT CGTGACTGGG AAACCCTGG CGTTACCCAA  
CAGCATAATG CCGCGGAGTG ACCGGCAGCA AAATGTTGCA GCACGTGACC TTTTGGGACC GCAATGGGTT

1331 CTTAATCGCC TTGCAGCACA TCCCCTTTC GCCAGCTGGC GTAAPAGCGA AGAGGCCCGC ACCGATCGCC  
GAATTAGCGG AACGTCGTGT AGGGGGAAG CCGTCCGACC CAATATCGCT TCTCCGGGCG TGGCTAGCGG

1401 CTTCCCAACA GTTGGCAGC CTGAATGGC AATGGGACG GCCCTGTAGC GGCCCATTA GCGCGGCGGG  
GAAGGGTGT CAACGCGTGG GACTTACC GC TACCCTGG CCGGACATCG CCGCGTAATT CCGCGCCGCC

1471 TGTGGTGGTT ACGCGCAGG TGAACCGTAC ACTTGGCCAG GCCCTAGCGC CCGTCCCTTT CGCTTCTTC  
ACACCACCAA TGGCGCTCC ACTGGCGATG TGAACGGTCC CCGGATCGCG GCGGAGGAAA GCGAAGGAG

1541 CCTTCCCTTC TCGCCACGTT CCGCGGCTTT CCGCGTCAAG CTCFAAATCG GGGGCTCCCT TTAGGGTTCC  
GGAAGGAAAG AGCGGTGCAA CCGGCCGAAA GGGGCAGTTC GAGATTTAGC CCCCAGGGA AATCCCAAG

1611 GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAACITGA TTAGGCTGAT GGTTCACGTA GTGGGCCATC  
CTAAATCAGC AATGCCGTG GAGCTGGGGT TTTTGTGACT AATCCACTA CCAAGTGCAT CACCCGGTAG

1681 GCCCTGATAG ACGGTTTTTC GCCCTTGGAC GTTGGAGTCC ACGTCTTTA ATAGTGGACT CTTGTTCCAA  
CGGGACTATC TGCCAAAAG CCGGAAACTG CAACCTCAGG TGCAGAAAT TATCACCTGA GAACAAGGTT

1751 ACTGGAACAA CACTCAACC TATCTGGTTC TATTCTTTG APTTAFAAG GATTTGCGG APTTCGGCCT  
TGACCTTGTG GTGAGTTGG ATAGAGCCAG ATAGAAAAC TAAATATTC CTAARACGGC TAAAGCCGGA

1821 ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTTAACG GAATTTAAC AAAATATTA CGCTTACAAT  
TAACCAATTT TTTACTCGAC TAAATGTGTT TTAATTTGCG CTTAAAATG TTTTATAATT GCGAATGTTA

FIG. 43D

1891 TTAGGTGGCA CTTTTCGGGG AAATGTGGC GGAACCCCTA TTGTTTATT TTTCTAAATA CATTCAAATA  
 AATCCACCGT GAAAAGCCCC TTACACGGG CCTTGGGGT AAACAATAA AAGATTTAT GTAAGTTTAT

1961 TGTATCCGCT CATGAGACAA TAACCCGTGAT AAATGCTTCA AFAATATTGA AAAAGGAAGA GTATGAGTAT  
 ACATAGCGGA GTACTCTGTT AFTGGGACTA TTTTACGAAGT TAFTATAACT TTTTCTTCT CATACTCATA

2031 TCAACATTTT CFTGTCCCC TTATTCCTTT TTTTGGGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA  
 AGTTGTAAG GCACAGCGG AATAAGGGA AAAACGCCGT AAAACGGAAG GACAAAACG AGTGGTCTT

2101 ACGTGGTGA AAGTAAAGA TGCTGAAGAT CAGTTGGGTG CACGAGTGG TTACATCGAA CTGGATCTCA  
 TGGGACCACT TTTCATTTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT

2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAANG ATGAGCACTT TTAAGTTCT  
 TGTCCGCCAT CTAGGAATC TCAAAAGCGG GBCTTCTTGC AAAAGTTAC TACTCTGTGA AATTTCAAGA

2241 GCTATGTGGC GCGGTATTTT CCCGTATTGA CGCCGGGCAA GAGCAACTCG GTCGCCGCAT ACACATTTCT  
 CGATACACCG CGCCATAATA GGCATAACT CGGCCCGT CTCGTTGAG CAGCGGCGTA TGTGATAAGA

2311 CAGAATGACT TGGTTGATA CTCACCAGTC ACAGAAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAT  
 GTCTTACTGA ACCAATCTAT GAGTGGTCAG TGTCTTTTTC TAGAATGCTT ACCGTACTGT CATTCCTTTA

2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACACTGCGGC CAACTTACTT CTGACAACGA TCGGAGGACC  
 ATACGTCACG ACGGTATTGG TACTCACTAT TGTGACGCCG GTTGAATGAA GACTGTGCT AGCCTCCTGG

2451 GAAGGAGCTA ACCGCTTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCCGGAG  
 CTTCCTCGAT TGGCGAAAAA ACGTGTGTA CCCCCTAGTA CATTGAGCGG AACTAGCAAC CCTTGGCCTC

2521 CTGAATGAAG CCAATACAAA CGACGAGCGT GACACCACGA TGCCTGTAGC AATGGCAACA ACGTTCCGCA  
 GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGTGGTGT ACGGACATCG TTACCGTTGT TGCACCCGCT

2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG AGCGGATAA  
 TTGATAATTG ACCGCTTGAT GATGAGATC GAAGGCCCT TGTAAATTAAT CTGACCTACC TCCGCCATT

2661 AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT  
 TCAACGTCCT GGTGAAGACG CGAGCCGGGA AGCCGACCG ACCAAATAAC GACTATTAG ACCTCGGCCA

FIG. 43E

2731 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGSCAG ATGGTAAGCC CTCCCCTATC GTAGTTATCT  
 CTCGCACCCA GAGCGCCATA GTAACGTCTG TACCCCGGTC TACCATTCCG GAGGGCATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGATAGGTG CCTCACCTGAT  
 TGTGCTGCC CTCACTCCGT TGATACCTAC TTGCTTTTATC TGCTTAGCGA CTCTATCCAC GGAGTGCATA

2871 TAAGCATGG TAACTGTGAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTTAA  
 ATTCGTAACT ATTGACAGTC TGGTTCAAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAATT

2941 TTTTAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAAT CCTTAAACGT GAGTTTCGT  
 AAATTTTCTT AGATCCACTT CTAGGAAAAA CTATTAGAT ACTGGTTTTA GGGAAATTGCA CTCAAAAAGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TCGGCCATAAT  
 AGGTGACTCG CAGTCTGGG CATCTTTTCT AGTTTCCCTAG AAGAACTCTA GAAAAAAAAG ACGGCAATTA

3081 CTGCTGCTTG CAAACAATAA AACCCCGCT ACCAGCGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT  
 GACGACGAAC GTTTGTTTTT TTGGTGGCGA TGGTCGCCAC CAAACAACG GCCTAGTCT CGATGGTTGA

3151 CTTTTTCCGA AGGTAACTGG CTTCAGCAGA GCGCAGATAC CAAAFACGTG CTTCTAGTG TAGCCGTAGT  
 GAAAAAGGCT TCCATFGACC GAACTCTCT GCGGCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCITACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
 ATCCGGTGGT GAAGTTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGGTCACCG

3291 TGCTGCCAGT GCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGGCCAG  
 ACGACGGTCA CCGCTATTCA GCACAGATG GCCCAACCTG AGTTCTGCTA TCAATGGCCT ATTCGGCGTC

3361 CGGTGCGGCT GAACGGGGG TTCGTGCACA CAGCCCAGCT TGGACGAAAC GACCTACACC GAACCTGAGT  
 GCCAGCCCGA CTTGCCCCCC AAGCACGTGT GTCGGGTGCA ACCTCGCTTG CTGGATGTGG CTTGACTCTA

3431 ACCTACAGCG TGAGCTATGA GAAAGCCCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAAG  
 TGGATGTCCG ACTCGATACT CTTTCGCGGT GCGAAGGGCT TCCCTCTTTC CGCCTGTCCA TAGGCCATTC

3501 CGGCAGGGTC GGAACAGGAG ACCGCACGAG GGAGCTTCCA GGGGAAAACG CCTGGTATCT TTATAGTCT  
 GCCGTCCCAG CCTTGTCTTC TCGGCTGCTC CCTCGAAGGT CCCCCTTTCG GGACCATAGA AATATCAGGA

FIG. 43F



3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTTGT GATGTCGT AGGGGGGCGG AGCCTATGGA  
CAGCCCAAG CGGTGGAGAC TGAACCTCGA GCTAAAACA CTACGAGCAG TCCCCCGGCC TCGGATACCT

3641 AAAACGCCAG CAACGGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCT TTGCTCACA TGTCTTTCC  
TTTTGCGGTC GTTGGCCGG AAAAATGCCA AGGACCGAA AACGACCGA AACGAGTGT ACAAGAAAGG

3711 TCGTTATCC CCTGATTCG TGGATAACCG TATTACGCC TTTGAGTGAG CTGATACCCG TCGCCCGCAG  
ACGCAATAGG GGAATAAGAC ACCTATTGGC ATAATGGCG AACTCACTC GACTATGGC AGCGGCGTCC

3781 CGAACGACC AGCGCAGCGA GTCAGTGAG GAGGAAGCG GAGAGGCC AATACGCAA CCGCCTCTCC  
GCTTGCTGGC TCGGCTCGT CAGTCACTC CTCCTTCGCC TTCTCGCGG TTATGCGTTT GCGGGAGAGG

3851 CCGCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTCCCGAC TGGAAAGCGG GCAGTGAGCG  
GGCGCGCAAC CCGCTAAGTA ATTACGTCGA CCGTGTGTC CAAAGGCTG ACCTTTCGCC CGTCACTCCG

3921 CAACGCAAT AATGTGAGT AGTCACTCA TTAGGCACC CAGGCTTAC ACTTATGCT TCCGGCTCGT  
GTTGCGTTAA TTACACTCAA TCGAGTGAGT AATCCGTGG GTCCGAAATG TGAANTACGA AGGCCGAGCA

3991 ATGTTGTGTG GAATTGTGAG CCGATAACAA TTTACACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC  
TACAACACAC CTTAACACTC GCCTATTGTT AAAGTGTGTC CTTTGTGAT ACTGGTACTA ATGCGGTTCCG

Bs8HII  
\*\*\*\*\*  
EcoRI  
\*\*\*\*\*

4061 GCGCAATTA CCTCACTAA AGGGAACAAA AGCTGGRATT CCACAATGAA CAATAATAAG ATTAAAATAG  
CGCGTTAAT GGGAGTGAT TCCCTTGTTT TCGACCTTAA GGTGTACTT GTTATTTATC TAATTTTATC

4131 CTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAAA TAAAAGATA ACTTAGACTC AAAACATTTA  
GAACGGGGC AACGTCGCTA CCCATAAAA AGATCATTTT ATTTCTATT TGAATCTGAG TTTTGTAAAT

4201 CAAAACAAC CCCTAAAGTC CTAAAGCCCA AAGTGTATG CAGATTCAT AGCAAGCCCA GCCCAACCCA  
GTTTTGTTG GGGATTTTCAG GATTTGGGT TTCACGATAC GTGCTAGTA TCGTTCGGT CGGGTTGGT

FIG.- 43G

4271 ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC CCGGCACAT CACCGTAGT CACCGTAGT  
 TGGGTTGGGT TGGGTTGGGT CACGTCGGTT GACCGTTAT CAGAGGTGG GGCCTGTGATA GTGGCACTCA

4341 TGTCGCACC ACCGCACGTC TCGCAGCCAA AAAAAAAAAA AGAAGAAAA AAAAGAAAA GAAAAACAGC  
 ACAGGCGTGG TGGCGTGCAG AGCGTCGGTT TTTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTTGTGG

4411 AGGTGGTCC GGTTCGTGG GCGCGGAAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC  
 TCCACCCAGG CCCAGCACCC CCGGCTTTT CCGTCTTCT AGCGTCTGC GTGCTCCGG GCCGGGAGGG

4481 TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATAATCA TACCCCTCCC TCTCTCCCA TCCCCCAAC  
 AGCGAAGGT TCTTTGCGG GGGGTAGCG TGATATATGT ATGGGGGGG AGAGGAGGGT AGGGGGTGG

4551 CCTACCACCA CCACCACCAC CACCTCTCC CCCCTCGCTG CCGGACGACG AGCTCTCCC CCTCTCCCCT  
 GGATGTTGGT GGTGTTGGTG GTGGAGGAG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG GGGAGGGGA

4621 CCGCCGCCG CCGTAACCAC CCGCCCTCT TCTCTTTCT TCTCTCTTT TTTTTTTCGT CTCGGTCTCG  
 GCGCGCGCG GCCATTGGTG GCGCGGGGAG AGGAGAAAAG AAGAGCATA AAAAAAGCA GAGCCAGAGC

4691 ATCTTTGGC TTGGTAGTT GGTGCGCGA GAGCGCTTC GTCGCCAGA TCGGTGCGG GAGGGGGGG  
 TAGAAACCGG AACCATCAA CCCACCCGT CTCGCCGAG CAGCGGTCT AGCCACCGC CCTCCCCGCC

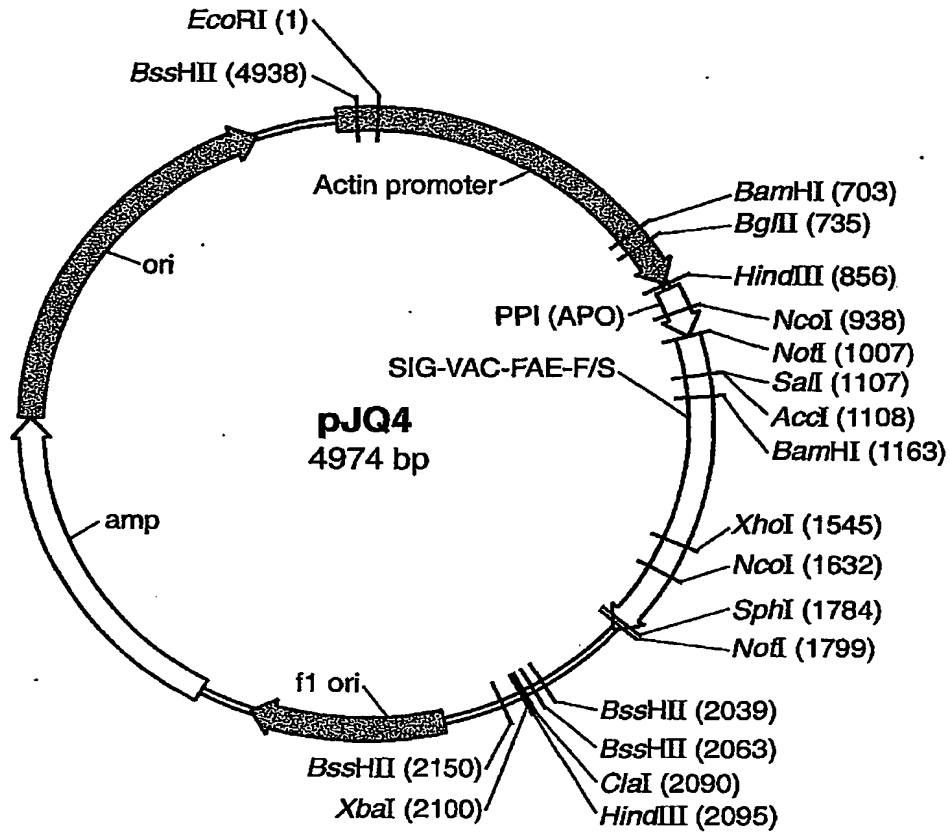
4761 GATCTCGCG CTGGCTCTC CCGGCGTGG TCGGCCGGA TCTCTCGGG GAATGGGGCT CTCGGATGTA  
 CTAGAGCGCC GACCCAGAG GCCCGCACTC AGCCGGGCTT AGGAGCGCC CTATACCCCGA GAGCCTACAT

BglII  
 BamHI

4831 GATCTCTTT CTTCTTCTT TTTGTTGTTG AATTGAAAT CCTCAGCAT GTTCATCGGT AGTTTTTCTT  
 CTAGAAGAAA GAAAGAGAA AACACCATC TTAACCTTAG GGAGTCTGTA CAAGTAGCCA TCAAAAAGAA

4901 TTCATGATTT GTGACAAATG CAGCTCTGT GCGAGCTTT TTGTAGGTAG  
 AAGTACTAAA CACTGTTTAC GTCGGAGCAC GCCTCGAAA AACATCCATC

FIG. 43H



**FIG. 44A**

**EcoRI**  
 ~~~~~  
 1 AATCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTTAGT  
 TTAAGGTGTT ACTTGTATT ATTCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA  
  
 71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTACAAAAA CAACCCCTAA AGTCTAAAG CCCAAGTGC  
 TTTTATTTTC TATTGAAATC TGAGTTTTGT AAATGTTTTT GTTGGGGGATT TCAGGATTTT GGGTTTCACG  
  
 141 TATGCACGAT CCATAGCAAG CCCAGCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA  
 ATACGTGCTA GGTATCGTTC GGGTCGGGT GGGTTGGGT GGGTCACGTC GGTTCACCGT  
  
 211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAAAA  
 TTATCAGAGG TGGGGCCCT GATAGTGCA CTCACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT  
  
 281 AAAAAGAAAG AAAAAAAGA AAAAANA CAGCAGGTGG GTCCGGGTCT TGGGGCCCG AAAAGCGAGG  
 TTTTCTTTTC TTTTTTTTCT TTTTCTTTT TTTTCTTTT GTCGTCCACC CAGGCCCAGC ACCCCCGCC TTTTTCGTCC  
  
 351 AGGATCGGA GCAGCGAGA GGCCTGGCC TCCCTCCGT TCCAAAGAAA CGCCCCCAT CGCCACTATA  
 TCCTAGCGCT CGTCTGTCT CCGGCGCGG AGGAGGCGA AGGTTTCTTT GCGGGGTA GCGGTGATAT  
  
 421 TACATACCC CCCCTCTCT GGTAGAGGA GGTAGAGGG GTTGGGATGG ACCACCACCA CCACCACCTC CTCCCCCCTC  
 ATGTATGGG GGGGAGAGG GGTAGAGGG GTTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGAG  
  
 491 GCTGCCGGAC GACGAGCTCC TCCCTCTCC CCGCGGTA CCGCGGTA CCAACCCGCT CCTCTCTCT  
 CGACGGCCTG CTGCTCGAG AGGGGGAGG GGGAGCGGC GCGGGCCATT GTGGGGCGG GGAAGGAGA  
  
 561 TTCTTCTCC GTTTTTTTT TCGTCTGGT CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG  
 AAGAAAGAGG CAAAAAAA AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC  
  
 631 CTTCGTGCC CAGATCGGTG CCGGGAGGG GCGGATCTC GCGGCTGGC TCCTCGGGCG TGAGTCGGCC  
 GAAGCAGCGG GTCTAGCCAC GCGCCCTCC CCGCCTAGAG CCGCCGACCG AGAGGCCCG ACTCAGCCCG  
  
  
**BamHI**  
 ~~~~~  
 701 CGGATCCTCG CGGGGAATGG GGCCTCGGA TGTAGATCTT CTTCTTTCT TCTTTTGTG GTAGAATTTG  
 GCCTAGGAGC GCCCCTTACC CCGAGAGCT ACATCTAGAA GAAAGAAAGA AAAAAACAC CATCTTAAAC

**FIG. 44B**

771 AATCCCTCAG CATTGTTCAAT CCGTAGTATT TCTTTTCATG APTTGTGACA AATGCAGCCT CGTGGGGAGC  
TTAGGGAGTC GTACACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACCGTCGGA GCACGCCCTCG

HindIII  
\*\*\*\*\*

841 TTTTTTGTAG GTAGAAGCTT ACMATGGMCG TGCACAAGGA GGTFAACTTC GTSGCCTACC TCCTGATCGT  
AAAAAACATC CATCTTCGAA TGTACTCKGC ACGTGTTCCT CCASTTGAAG CASC GGATGG AGGACTAGCA

NcoI  
\*\*\*\*\*

911 SCTCGGCCTC CTCITGCTCG TSTCCGCCAT GGAGCAGTG GACGCCAAG CCTGCACCKK CGAGTGGGGC  
SGAGCCGGAG GAGAACGAGC ASAGCCGGTA CCTCGTGCAC CTGCGTTC GACCGTGGM GCTCACGCCCG

NotI  
\*\*\*\*\*

981 AACCTCGGCT TCGGCATCTG CCCGGGGCC CCTCCACGC AGGCAATTC CGAAGACCTC TACAGCCGTT  
TTGGAGCCGA AGCCGTAGAC GGGCCGCCGG CGAGGTGG TCCCSTAGAG GCTTCTGGAG ATGTCGGCAA

Sall  
\*\*\*\*\*

AcoI  
\*\*\*\*\*

1051 TAGTCGAAAT GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC APTCCGTCGA CTATTATCAA  
ATCAGCTTTA CCGSTGATAG AGGTTGAC AGGTTGCGAC GGATGCGGCT GGACACGTTG TAAGGCAGCT GATPATAGTT

BamHI  
\*\*\*\*\*

1121 GGGAGAGAAA ATTTACAAT CTCAAACTGA CATTACGGA CATTACGGA TGGATCCYCC GCGACGACAG CAGCAAAGAA  
CCCTCTCTTT TAAATGTTAA GAGTTGACT GTAATTGCCT ACCTAGGAGG CGCTGCTGTC GTCGTTTCTT

1191 ATAATCACCG TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAACTAC ACCCTCACGC  
TATTAGTGGC AGAAGGCACC GTGACCAATCA CTATGCTTAG ATGTTGAGCT ATGATTTGATG TGGGAGTGGC

1261 CTTTCGACAC CCTACCACAA TGCACCGGTT GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT  
GAAAGCTGTG GGATGGTGTG ACGTTCGCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA

FIG.-44C

1331 CCAGGACCACAA GTCGAGTCCG TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGGCTT GACCGTGACC  
GGTCCCTGGTT CAGCTCAGCG AACAGTTTGT CGTCCAATCG GTCATAGGCC TGATGCGCGA CTGGCACATGG

1401 GGCACCKCC TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGGACATAC GACAACATCC  
CCGGTGMGGG AGCCGCGGAG GEAACCGCGT GAGTGACGGC GGGTCGACAG ACGCTGTATG CTGTTGTAGG

1471 GCCTGTACAC CTTCCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCCGGCTG TACATGAACG ATGCCTTCCA  
CGGACATGTG GAAGCCGCCTT GCGCGCTCG CGTTAGTCCG GAAGCGCAGC ATGTACTTGC TACGGGAAGGT

XhoI  
\*\*\*\*\*

1541 AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC  
TCGGAGCTCG GGTCTATGCT GCTCATATAA GCGCCAGTGA GTACGGTTGC TGCCGTAGGG TTTGGACGGG

NcoI  
\*\*\*\*\*

1611 CCGGTGGAGC AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAAC  
GGCCACCTCG TCCCCATGCG GGTACCGCCA CATCTCATGA CCTCGCAACT AGGAATGTCC CGGTCITGT

1681 CATTTGTCG CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA  
GTAAACAGAC GTGACCCCTA CTTACAGTCA CGACACTCCG GGTCCCGCCT GTCCCACACT TATTACGGCT

SphI  
\*\*\*\*\*

NotI  
\*\*\*\*\*

1751 CACGACTTAT TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTGCGCG CCGCGGAAC CACTGAAGGA  
GTGCTGAATA AAACCCCTACT GCTCCGCGCG TACGTGGACC GGCCAGCGCC GCGCCTTTG GTGACTTCTT

1821 TGAGCTGTAA AGAAGCAGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC CTGTTGCCGG  
ACTCGACATF TCTTTCGTCTA GCAAGTTTGT AAACCGTTAT TTCAAAGAAAT TCTAACTTAG GACAACGGCC

1891 TCTTGGGATG ATTATCATAT AATTTCTGTT GAATTACGTT AAGCATGTA TAATAACAT GTAATGCATG  
AGAACGCTAC TAATAGTATA TTAAGACAA CTTAATGCAA TTCGTACATF ATTAATGTA CATTACGTAC

1961 ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCC GC AATTAATACAT TTAATACGG ATAGAAAACA  
TGCAATAAAT ACTCTACCCA AAAATACTAA TCTCAGGGCG TTAATAATGTA AATTATGCC TATCTTTTGT

FIG.- 44D



2731 TTTAAGCGA ATTTAACA AATAAATACG CTTACAATTT AGGTGGCACT TTTGGGGAA ATGTGGCGG  
 AAATTGGCT TAAAATTGTT TTAFAATGC GAATGTARA TCCACCGTGA AAAGCCCTT TACACGGCC

2801 AACCCCTATT TGTTTATTTT TCTAATACA TTTCAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA  
 TTGGGGATAA ACAAATAAAA AGATTTATGT AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGGACTATT

2871 ATGCTTCAAT AATATTGAA AAGGAAGAT ATGAGTATTC AACATTTCCG TGTGCCCCCTT ATTCCTTTT  
 TACGAAGTTA TTATAACTTT TTCCTTCTCA TACTCATAAG TTGTAAGGC ACAGCGGAA TAAGGGAATA

2941 TTGCGGCATT TTGCCTTCTT GTTTTGTCTC ACCCAGAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA  
 AACGCCGTAA AACGGAAGGA CAAAACGAG TGGGTCTTTG CGACCCTTT CATTTCCTAC GACTTCTAGT

3011 GTTGGGTGCA CGAGTGGGT ACATCGAAT GGATCTCAAC AGCGTAAGA TCCTTGAGAG TTTTCGCCCC  
 CAACCCACGT GCTCACCCAA TGTAGCTGA CCTAGAGTTG TCGCCATTCT AGGAACCTC AAAAGCGGGG

3081 GAAGAAGCTT TTCCAATGAT GAGCACTTTT AAAGTCTGC TATGTGGCG GGTATATATCC CGTATTGACG  
 CTTCTTGCAA AAGTTACTA CTCGTGAAA TTTCAAGACG ATACACCCGC CCATAATAGG GCATTAACCTGC

3151 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATCTCA GAATGACTTG GTTGAGTACT CACCAGTCA  
 GCGCCGTTCT CGTTGAGCCA GCGCGTAG TGATAAGAGT CTTACTGAA CAACTCATGA GTGGTCAATG

3221 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATA TGCAGTGTG CCATAACCAT GAGTGATAAC  
 TCTTTTCGTA GAATGCCATC CGTACTGTCA TTCTCTTAT ACGTACGAC GGTATTTGTA CTCACTATTG

3291 ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGTAA CCGTTTTTTG CACAACATGG  
 TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAC GTGTTGTACC

3361 GGGATCATGT AACTCGCCTT GATCGTTGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA  
 CCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCCG TATGGTTTGC TGCTCGCACT

3431 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGGCMAA CTATTAATCG GCGAARTACT TACTCTAGCT  
 GTGTGCTTAC GGACATCGTT ACCGTGTTG CAACCGCTTT GATAATTGAC CGCTTGATGA ATGAGATCGA

3501 TCCCGGCAAC AATTAATAGA CTGATGGAG GCGGATAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC  
 AGGCCGTTG TTAATTATCT GACCTACCTC CGCTATTTC AACGTCTTGG TGAAGACGG AGCCGGGAG

FIG. 44F



3571 CCGCTGGCTG GTTTATTGGT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT  
 GCCGACCCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT AACGTCTGTGA

3641 GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGA GTCAGGCAAC TATGGATGAA  
 CCCCGGTCTA CCATTCGGGA GGGCATAGCA TCAATAGATG TGCTGCCCTT CAGTCCGTTG ATACCTACTT

3711 CGAAATAGAC AGATCGCTGA GATAGGNSCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT  
 GCTTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAATGA

3781 CATATATACT TTAGATTGAT TTAARACTTC ATTTTAAAT TAAAGGATC TAGGTGAAGA TCCTTTTGA  
 GTATATATGA AATCTAACTA AATTTTGAAG TAAARATTA ATTTTCCCTAG ATCCACTTCT AGGAAAAACT

3851 TAATCTCATG ACCAAAATCC CTTAACGTTGA GTTTTCGTTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC  
 APTAGAGTAC TGGTTTTAGG GAATFGCACT CAAAAGCAAG GTGACTCGCA GTCGTGGGCA TCTTTTCTAG

3921 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CCGTAATCT GCTGCTTGA AACAAAAAA CCACCCGCTAC  
 TTTCTTAGRA GAATCTTAGG AAAAAAGAC GCGCATTAGA CGACGAACTG TTGTTTTTTT GGTGGCGATG

3991 CAGCGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAG GTAACCTGGT TCAGCAGAGC  
 GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTGACCGA AGTCGTCTCG

4061 GCAGATACCA AATACTGTCC TTCTAGTGA GCCGTAGTTA GGCCACCACT TCAAGACTC TGTAGCACCG  
 CGTCTATGGT TTATGACAGG AAGATCACAT CCGCATCAAT CCGGTGGTGA AGTTCCTGAG ACATCGTGGC

4131 CCTACATACC TCGCTCTGCT AATCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCC TGTCTTACC  
 GGATGTATGG AGCGAGACGA TTAGACAAT GGTACCCGAC GACGGTCACT GCTATTCAGC ACAGAATGGC

4201 GGTGGACTC AAGACGATAG TTACCAGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCCACACA  
 CCAACCTGAG TTCTGCTATC AATGGCTAT TCCGGTCCG CAGCCCGACT TGCCCCCAA GCACGTGTGT

4271 GCCCAGCTTG GAGGGAACGA CCTACACCGA ACTGAGATAC CTACAGCTG AGCTATGAGA AAGCGCCACG  
 CCGGTCCGAC CTCGGCTTGT GGAATFGGCT TGACTCTATG GATGTCGCAC TCGATACTCT TTCGCGGTGC

4341 CTTCCCCGAG GGAGAAAGC GGACAGGTAT CCGGTAAGCG GCAGGGTCCG AACAGAGAG CGCACGAGGG  
 GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCGC CGTCCCAGCC TTGTCTCTTC GCGTGTCTCC

FIG. 44G

4411 AGCTTCCAGG GGAAGAGCC TGGTATCTTT ATAGTCCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCCGTCG  
TCGAAGGTCC CCTTTGCGG ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC

4481 ATTTTGTGA TGTCTGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGFTTC  
TAAAAACACT ACGAGCAGTC CCCCAGCCTC GGATACCTTT TTGCGGTCTGT TCGCCCGGAA AATGCCCAAG

4551 CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTTCTTCCCTG CATTATCCCC TGTTCCTGTG GATAACCCGTA  
GACCGGAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT

4621 TTACCGCCTT TGAGTGAAGT GATACCCTC GATACCCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAAGCA  
AATGGCGGAA ACTCACTCGA CTATGGCGAG CCGCGTCCGC TTGCTGGCTC GCCTCGCTCA GTCACCTCGCT

4691 GGAAGCGGAA GAGGCCCAA TACGCAAACC GCCTCTCCC GCGCTTGGC CGATTCAATTA ATGCAGCTGG  
CCTTCGCCCTT CTCGCGGCTT ATCGTCTGG CCGAGAGGGG CCGCAACCG GCTAAGTAAT TACGTCGACC

4761 CACGACAGGT TTCCCAGCTG GAAAGCGGGC AGTGAAGCCA ACGCAATTAA TGTGAGTTAG CTCACCTCAT  
GTGCTGTCCA AAGGCTGAC CTTTCGCCCG TCACTCGCGT TGCCTTAATT ACACTCAATC GAGTGAATTA

4831 AGGCACCCCA GGCTTACAC TTTATGCTTC CCGCTCGTAT GTTGTGTGA ATTGTGAGCG GATAACAATT  
TCCGTGGGGT CCGAATGTG AAATACGAAG GCCGAGCATA CAACACACT TAACACTCGC CTATTGTATA

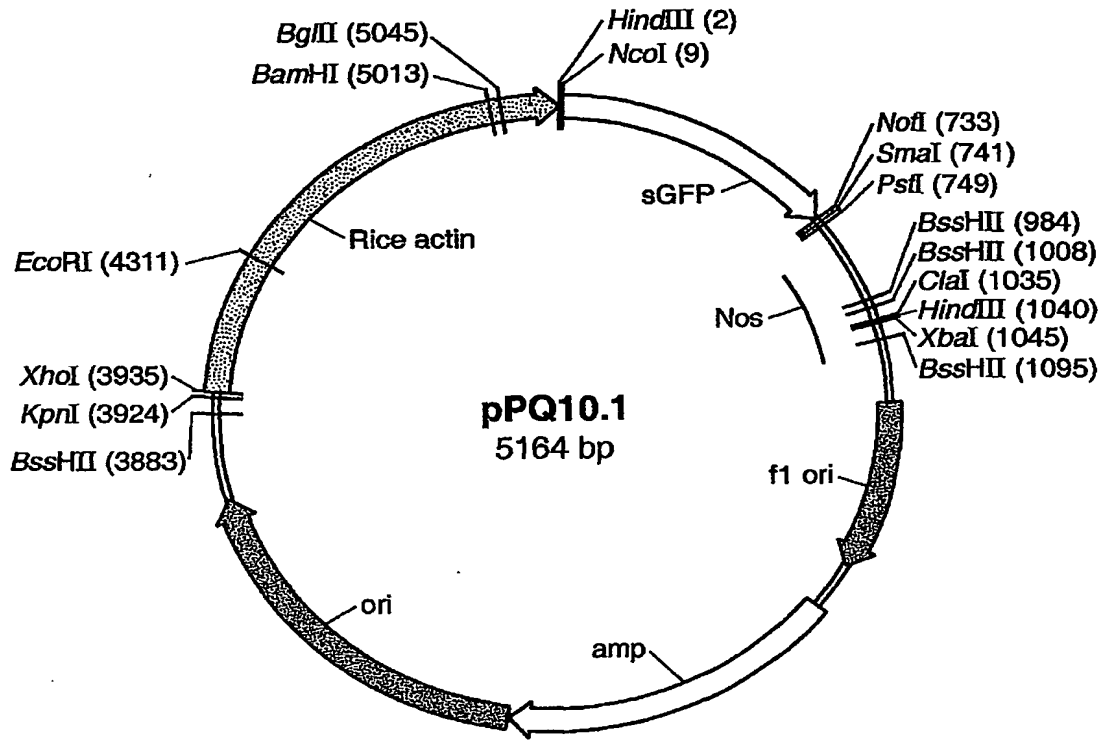
4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAGCGC GCAATTAACC CTCACTAAG GGAACAAAAG  
AGTGTCTCCT TTGTGATAC TGGTACTAAT GCGGTTCCG CATTAAATGG GAGTGAATTC CCTTGTCTTC

BSSHII  
\*\*\*\*\*

EcoR

CTGG  
GACC

FIG. 44H



**FIG. 45A**

## Sequence for pPQ10.1

HindIII NcoI

```

-----
1  AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCCG GGGTGGTGCC
CATCCTGGTC GAGCTGGACG
      TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG
GTAGGACCAG CTCGACCTGC

71  GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT
GCCACCTACG GCAAGCTGAC
      CGCTGCACTT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA
CGGTGGATGC CGTTCGACTG

141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCCTGCCC TGGCCCACCC
TCGTGACCAC CTTCACCTAC
      GGACTTCAAG TAGACGTGGT GGCCTGTCGA CGGGCACGGG ACCGGGTGGG
AGCACTGGTG GAAGTGGATG

211 GCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGGACTT
CTTCAAGTCC GCCATGCCCG
      CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTGAA
GAAGTTCAGG CGGTACGGGC

281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC
AAGACCCGCG CCGAGGTGAA
      TTCCGATGCA GGTCTTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG
TTCTGGGGCG GGCTCCACTT

351 GTTCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT
TCAAGGAGGA CGGCAACATC
      CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA
AGTTCCTCCT GCCGTTGTAG

421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT
GGCCGACAAG CAGAAGAACG
      GACCCCGTGT TCGACCTCAT GTTGATGTG TCGTGTGTC AGATATAGTA
CCGGCTGTTT GTCTTCTTGC

491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG
CAGCTCGCCG ACCACTACCA
      CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTGCGAC
GTCGAGCGGC TGGTGATGGT

561 GCAGAACACC CCCATCGGCG ACGGCCCGT GCTGCTGCCC GACAACCACT
ACCTGAGCAC CCAGTCCGCC
      CGTCTTGTGG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA
TGGACTCGTG GGTGAGGCGG

631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT
CGTGACCGCC GCCTGGATCA

```

Fig. 45B

CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCAACGCG  
TCGGACTTAC CGCTTACCCT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGGGC GGGTGTGGTG GTTACGCGCA  
CGGTGACCGC TACACTTGCC  
GCGCGGGACA TCGCCGCGTA ATTGCGCCG CCCACACCAC CAATGCGCGT  
CGCACTGGCG ATGTGAACGG

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC  
GTTGCGCCGC TTTCCTCCGTC  
TCGCGGGATC GCGGGCGAGG AAAGCGAAAG AAGGGAAGGA AAGAGCGGTG  
CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG  
CACCTCGACC CCAAAAACT  
TTCGAGATTT AGCCCCGAG GGAATCCCA AGGCTAAATC ACGAAATGCC  
GTGGAGCTGG GGTFTTTTGA

1471 TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT  
TTCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAAA  
AAGCGGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTT CAAACTGGAA CAACACTCAA  
CCCTATCTCG GTCTATTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT  
GGGATAGAGC CAGATAAGAA

1611 TTGATTTATA AGGGATTTTG CCGATTTCCG CCTATTGGTT AAAAAATGAG  
CTGATTTAAC AAAAATTTAA  
AACTAAATAT TCCCTAAAAC GGCTAAAGCC GGATAACCAA TTTTTFACTC  
GACTAAATTG TTTTAAATT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTTAGGTG GCACFTTTTCG  
GGGAAATGTG CGCGGAACCC  
GCGCTTAAAA TTGTTTTATA ATTGCGAATG TTAAATCCAC CGTGAAAAGC  
CCCTTIACAC GCGCCTTGGG

1751 CTATTTGTTT ATTTTTCTAA ATACATTCOA ATATGTATCC GTCATGAGA  
CAATAACCCT GATAAATGCT  
GATAAACAAA TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT  
GTTATTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCTG  
CCCTTATTCC CTTTTTTGCG  
AGTTATTATA ACTTTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC  
GGGAATAAGG GAAAAACGC

1891 GCATTTTGCC TTCTGTFTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA  
AGATGCTGAA GATCAGTTGG  
CGTAAACGG AAGGACAAAA ACGAGTGGGT CTTTGCGACC ACTTTCATTT  
TCTACGACTT CTAGTCAACC

Fig. 45C

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT  
GAGAGTTTTT CCCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCGC ATTCTAGGAA  
CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTTCCA ATGATGAGCA CTTTTAAAGT TCTGCTATGT GGCGCGGTAT  
TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAAATTTCA AGACGATACA CGCGGCCATA  
ATAGGGCATA ACTGCGGCC

2101 CAAGAGCAAC TCGGTGCGCC CATACTACTAT TCTCAGAATG ACTTGGTTGA  
GTACTCACCA GTCACAGAAA  
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT  
CATGAGTGGT CAGTGTCTTT

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATATATGCAG TGCTGCCATA  
ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCAITCTC TTAATACGTC ACGACGGTAT  
TGGTACTCAC TATTGTGACG

2241 GGCCAACCTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT  
TTTTGCACAA CATGGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCTC GATTGGCGAA  
AAAACGTGTT GTACCCCTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC  
AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAGTAGC AACCTTGGC CTCGACTTAC TTCGGTATGG  
TTTGCTGCTC GCACTGTGGT

2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACCTATT AACTGGCGAA  
CTACTTACTC TAGCTTCCCG  
GCTACGGACA TCGTTACCGT TGTGCAACG CGTTTGATAA TTGACCGCTT  
GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC  
TGCGCTCGGC CFTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCT ATTTCAACGT CCTGGTGAAG  
ACGCGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG  
TATCATGCA GCACGCGGC  
CCGACCAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC  
ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG  
GCAACTATGG ATGAACGAAA  
GTCTACCATT CGGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC  
CGTTGATAAC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT  
CAGACCAAGT TTACTCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA  
GTCTGGTTCA AATGAGTATA

Fig. 45D

2731 ATACCTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTGATTAATC  
 TAGAATCT AACTAAATTT TGAAGTAAAA ATTAATTTT CCTAGATCCA CTCTAGGAA AACTATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CFTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG  
 AGTACTGGTT TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTCTTGA GATCCCTTTT TTCTGGCGT AATCTGCTGC TTGCAACAA AAAAAACCACC GCTACCAGCG  
 TAGAAGAACT CTAGGAAAAA AAGACGGCA TTAGACGACG AACGTTTGT TTTTTTGGTGG CGATGGTGGC

2941 GTGGTTTGT TTCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
 CACCAACAA ACGGCCTAGT TCTCGATGET TGAGAAAAAG GCTTCCATTG ACCGAAATCG TCTCGCGTCT

3011 TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTGTAG AACTCTGTAG CACCGCCTAC  
 ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGCTGCTGCC AGTGGCGATA AGTCTGTCT TACCGGGTTG  
 TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GNTAGTTACC GGATAAGCCG CAGCGTCCG GCTGAACGGG GGGTTCGTGC ACACAGCCCA  
 CTGAGTCTG CTATCAATGG CCTATTCCGC GTCCGACGCC CGACTTGCCC CCCAAGCACG TGTGTCTGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
 CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTTCG GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGCAGG GTCGGAACAG GAGAGCCAC GAGGGAGCTT  
 GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGT CTTCTCGCGTG CTCCCTCGAA

3361 CCAGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCTGGGT TTCGCCACCT CTGACTTGG CGTCGATTTT  
 GGTCCCTCTT TCGGGACCAT AGAATATCA GGACAGCCCA AAGCGGTGGA GACTGAATC GCAGTAAA

3431 TGTGATGCTC GTCAGGGGG CCGAGCCYAT GGAANAACGC CAGCAACGG GCTTTTATC GGTTCCTGGC  
 AACTACGAG CAGTCCCCC GCTTCCGATA CCTTTTGGG GTCGTTGGC CGGAAAAATG CCAAGGACCG

FIG. 45E

```

3501 CTTTTGCTGG CTTTTTGCCTC ACATGTTCTTT TCCTGCGGTTA TCCCTGATTT CTGTGGATAA CCGTATTACC
    GAAAACGACC GAAAACCGAG TGTACAAGAA AGGACGCAAT AGGSGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGGCGAG CGAGTCAGTG AGCGAGGAAG
    CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGGGTC GCTCAGTCAC TCGCTCCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCGGCGG TTGGCCGANT CATTAATGCA GCTGGCACGA
    GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGGCGGCG AACCGCTAA GTAATTACGT CGACCCGTGCT

3711 CAGGTTTCCC GACTGGAAG CCGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTTAGGCA
    GTCCAAAGGG CTGACCTTTT GCCCGTCACT CGCGTTCGGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATGT GAGCGGATAA CAATTTTACA
    GGGTCCGAA ATGTGAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT GTTAAAGTGT

          BssHII
          *****
3851 CAGGAACAG CTATGACCAT GATFACGCCA AGCGGCAAT TAACCTCAC TAAAGGGAAC AAAAGCTGGG
    GTCCTTTGTC GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGAATG ATTTCCCTTG TTTTTCGACCC

          KpnI
          *****
3921 TACCGGGCCC CCCCTCGAGG TCATTCATAT GCTTGAGAAG AGAGTCGGGA TAGTCCAAA TAAAACAAAG
    ATGGCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTTGTTC

          KpnI
          *****
3991 GTAAGATTAC CTGGTCAAAA GTGAAAACAT CAGTTAAAAG GTGGTAAAG TAAAATATCG GTARTAAAAG
    CATTCATATG GACCAGTTTT CACTTTTGTGTA GTCAATTTTC CACCATATC ATTTTATAGC CATTTATTTTC

4061 GTGGCCCAA GTGMAATTA CTCCTTTCTA CTATTATAA AATTGAGGAT GTTTGTGTCGG TACTTTGATA
    CACCGGGTTT CACTTTAAAT GAGAAAGAT GATAATATTT TTAACCTCTA CAAAACAGCC ATGAAACTAT
    
```

FIG. 45F



4131 CGTCATTTTT GTATGAATG GTTTTTAAGT TTATTTCGCGA TTGGAAATG CATATCTGTA TTGAGTCCGG TTGAGTCCGG  
GCAGTAAAA CAFACTTAAC CAAAAATCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACCTCAGCC

4201 TTTTTAAGTT CGTTGCTTTT GTAATAACAG AGGATTTGT AFAAGAAATA TCTTTAAAA ACCCATATGC  
AAAAATTCAA GCAACGAAAA CATTATATGTC TCCCPAAAA TATTCCTTTAT AGAAAATTTTT TGGGTATACG

EORI  
\*\*\*\*\*

4271 TAATTTGACA TAATTTTTGA GAAAAATATA TATTCAGCGG AATCCACAA TGAACAATAA TAAGATTAAA  
ATTAAACTGT ATTAAAACT CTTTTTATAT ATAAGTCCGC TTAAGGTGTT ACTTGTATT ATTCTAATTT

4341 ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTTCTAGT AAAATAAAG ATAACTTAG ACTCAAAACA  
TATCGAACGG GGGCAACGTC GCTACCATA AAAAGATCA TTTTATTTTC TATTTGAATC TGAGTTTTGT

4411 TTTACAAAA CAACCCCTAA AGTCTTAAAG CCCAAAGTGC TATGCACGAT CCATAGCAAG CCCAGCCCAA  
AAATGTTTT GTTGGGATTC TCAGGATTC GGGTTTCACG ATACGTGTA GGTATCGTTC GGGTCGGGTT

4481 CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCGGCA CTATCACCGT  
GGGTTGGGTT GGGTTGGGTT GGGTCACGTC GGTGACCGT TATCAGAGG TGGGGCCCGT GATAGTGGCA

4551 GAGTTGTCCG CACCACCGCA CGTCTGCAG CCAAAAAA AAAAAAAG AAAAAAAGA AAAAGAAAA  
CTCAACAGGC GTGTTGCCGT GCAGAGCGTC GGTTTTTTTT TTTTCTTTC TTTTCTTTT

4621 CAGCAGGTGG GTCCGGGTGG TGGGGCCGG AAAAGCGAGG AGGATCGCGA GCAGCGACGA GGGCCGGCCC  
GTCGTCCACC CAGGCCACG ACCCCGGCC TTTTCGCTCC TCCTAGCGT CGTCTGCTCT CCGGGCCGGG

4691 TCCCTCCGCT TCCAAAGAA CGCCCCCAT CGCCACTATA TACATACCCC CCCCCTCCT CCCATCCCCC  
AGGAGGCCGA AGGTTCTTTT CCGGGGGTA CCGGTGATAT ATGTATGGG GGGGAGGA GGGTAGGGGG

4761 CAACCTACC ACCACACCA CCACCACCTC CTCCCCCTC GCTGCCGAC GACGAGCTCC TCCCCCTCC  
GTTTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGGAG CGACGGCTG CTGCTCGAG AGGGGGGAGG

4831 CCCTCCGCG CCGCCGGTAA CCACCCGCG CCTCTCCTCT TTTCTTCTCC GTTTTTTTTT TCGTCTCGGT  
GGGAGCGCG GCGGCCATT GGTGGGCGG GGAGAGGAGA AAGAAGAGG CAAAAAAA AGCAGAGCCA

FIG. 45G

4901 CTCGATCTTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCTGTCGCC CAGATCGGTTG CCGGGGAGGG  
 GAGCTAGAAA CCGGAACCAT CAATCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC

BamHI

GGATCC

4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGATCCTCG CCGGGAATGG GGCTCTCGGA  
 CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG GCCTAGGAGC GCCCCCTACC CCGAGAGCCT

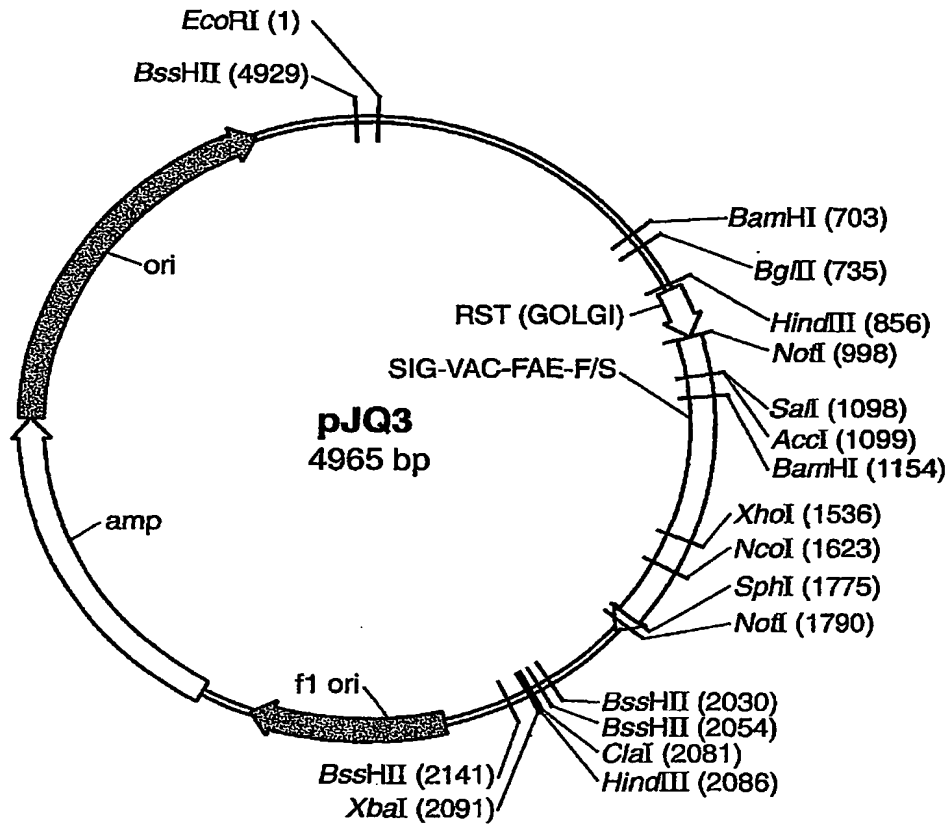
BglII

AGATCT

5041 TGATGATCTTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATTGTTCAT CGGTAGTTTT  
 ACATCTAGAA GAAAGAAGA AGAAAAACAC CATCTTAAC TTAGGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTCCGGAGC TTTTGTGTAG GTAG  
 AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG AAAAAACATC CATC

FIG. 45H



**FIG. 46A**

**EcoRI**  
 \* \* \* \* \*  
 1 AATCCACAA TGAACAATAA TAAGATATAA ATAGCTGGCC CCCGTTGCAG CGATGGGTAT TTTTCTAGT  
 TTAGGTGTT ACTTGTATTT ATTCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA  
  
 71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTTACAATAA CAACCCCTAA AGTCTAAAG CCCAAAGTGC  
 TTTTATTTTC TATTTGAATC TGAGTTTGT AAATGTTTTT GTTGGGATTC CAGGATTC GGGTTTCACG  
  
 141 TATGCACGAT CCAATGCAAG CCCAGCCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA  
 ATACGTGCTA GGTATCGTTC GGGTCGGGTT GGGTTGGGTT GGGTTGGGTT GGGTCAACGTC GGTTCACCGT  
  
 211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCACGCA CGTCTCGCAG CCAAAAAAAA  
 TTATCAGAGG TGGGGCCCGT GATAGTGGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT  
  
 281 AAAAAGAAG AAAAAAAGA AAAAAAATAA CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG  
 TTTTCTTTTC TTTTTTTTCT TTTTCTTTT TCTGTCCACC CAGGCCAGC ACCCCCGGCC TTTTTCGCTCC  
  
 351 AGGATCGCGA GCAGCGACGA GGGCCGGCCC TCCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCACTATA  
 TCCTAGCGCT CGTCGCTGCT CCGGGCCGGG AGGGAGCGA AGTTTCTTTT GCGGGGGGTA GCGGTGATAT  
  
 421 TACATACCC CCCCTCTCT CCCATCCCC CAACCTACC ACCACACCA CCACCACCCTC CTCCTCCCTC  
 ATGTATGGGG GGGGAGAGGA GGTAGGGGG GTTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGGAG  
  
 491 GCTGCCGGAC GACGAGCTCC TCCCTCCGCG CCCTCCGCGG CCGCCGGTAA CCACCCCGCC CCTCTCCTCT  
 CGACGGCCTG CTGCTCGAGG AGGGGGGAGG GGGAGGCGGC GCGGCGCAAT GGTGGGGCGG GGAGAGGAGA  
  
 561 TTCTTCTTCC GTTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGATA GTTTGGGTGG GCGAGAGCGG  
 AAGAAAGAGG CAAAAAATAA AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC  
  
 631 CTTCTGTCGC CAGATCGGTG CCGGGAGGG GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC  
 GAAGCAGCGG GTCTAGCCAC GCGCCTCCC CCGCCTAGAG CCGCCAGCCG AGAGGCCCGC ACTCAGCCCG  
  
**BamHI**  
 \* \* \* \* \*  
 701 CCGATCCTCG CGGGGAATGG GGCCTCGGA TGATAGTCTT CTTTCTTCT TCTTTTGTG GTAGAATTTG  
 GCCTAGGAGC GCCCCTTACC CCGAGAGCT ACATCTAGA GAAAGAAAGA AAAAAAACAC CATCTTAAC

FIG. 46B

771 AATCCCTCAG CATTGTTTCAAT CGGTAGTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGCGGAGC  
 TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCTCG

HindIII  
 ~~~~~

841 TTTTGTGTAG GTAGAAGCTT ACCATGATCC ACACCAACCT CAAAAGAAG TTCTCCCTCT TCATCTCTCGT  
 AAAAAACATC CATCTTCGAA TGGTACTAGG TGTGGTTGGA GTTTTCTTTC AAGAGGGAGA AGTAGGAGCA

911 CTTCTCCTC TTCGCCGTGA TCTGCTGTG GAAGAAGGC TCCGACTACG AGGCCCTCAC CCTCCAAGCC  
 GAAGGAGGAG AAGCGGCACT AGACGCACAC CTTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCCG

NotI  
 ~~~~~

981 AAGGAGTTCC AATGGCGGC CGCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCGT TTAGTCGAAA  
 TTCTCAAGG TTTACCGCCG GCGGAGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTTT

Sall  
 ~~~~~  
 AccI  
 ~~~~~

1051 TGGCCACTAT CTCCCAAGCT GCCTACGCCG ACCTGTGCAA CATTCCGTCG ACTATATATCA AGGGAGAGAA  
 ACCGGTGATA GAGGGTTCGA CCGATGCGGC TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

BamHI  
 ~~~~~

1121 AATTTACAAT TCTCAAAC TG ACATTAACGG ATGGATCCTC CGCGACGACA GCAGCAAAGA AATTAATCACC  
 TTAATGTTA AGAGTTTGAC TGTAAATTGCC TACCTAGGAG GCGCTGCTGT CGTCTGTTCT TTATTAGTGG

1191 GTCCTCCGTG GCACHTGGTAG TGATACGAAT CTACAACCTCG ATACTAACTA CACCCCTCAC CCTTTCCGACA  
 CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATTTGAT GTGGGAGTGC GGAAAGCTGT

1261 CCCTACCACA ATGCAACGGT TGTGAAGTAC ACGGTGGATA TTAATATGGA TGGGTCTCCG TCCAGGACCA  
 GGGATGGTGT TACGTTGCCA ACACTTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCTCTGGT

FIG.-46C

1331 AGTCGAGTCG CTTGTCAAAC AGCAGGTTAG CCAGTATCCG GACTACGCCG TGACCGTGAC CGGCCACKCC  
 TCAGCTCAGC GAACAGTTTG TCGTCCAATC GGTCATAGGC CTGATGCGCG ACTGGCACAG GCCGPTMGG

1401 CTCGGCGCCT CCCTGGCGGC ACTCACTGCC GCCCAGCTGT CTGGACATA CGACAACATC CGCCTGTACA  
 GAGCCGCGGA GGGACCGCCG TGAGTGACCG CGGGTCGACA GACGCTGTAT GCTGTTGTAG GCGGACATGT

XhoI  
 ~~~~~

1471 CCTTCGGCGA ACCGCGCAGC GGCAATCAGG CCTTCGCCCTC GTACATGAAC GATGCCCTTCC AAGCCTCGAG  
 GGAAGCCGCT TGGCGCGTCC CCGTTAGTCC GGAAGCCGAG CATGTACTTG CTACGGAAGG TTCGGAGCTC

1541 CCCAGATACG ACCCAGTATT TCCGGGTAC TCATGCCAAC GACGGCATCC CAAACCTGCC CCGGTTGGAG  
 GGGTCTATGC TCGGTCATAA AGGCCCAGTG AGTACGGTTG CTGCCGTAGG GTTTGGACGG GGGCCACCTC

NcoI  
 ~~~~~

1611 CAGGGTACG CCCATGGCGG TGTAGAGTAC TGGAGCGTTG ATCCTTACAG CGCCCAGAAC ACAATTGTCT  
 GTCCCATGC GGGTACCGCC ACATCTCATG ACCTCGCAAC TAGGAATGTC GCGGGTCTTG TGTAAACAGA

1681 GCACGGGGA TGAAGTCCAG TGCCTGTGAG CCCAGGGCGG ACAGGGTGTG AATAATGCCG ACACGACTTA  
 CBTGACCCCT ACTTCACGTC ACGACACTCC GGGTCCCGCC TGTCCACAC TTATTACCGG TGTGCTGAAT

SphI  
 ~~~~~

1751 TTTTGGGATG ACGAGCGCGG CATGCACCCTG GCCGGTCCGG GCCGGGAAA CCACGTGAAG ATGAGCTGTA  
 AAAACCTTAC TGCTCGCCGC GTACGTGGAC CGGCCAGCCG CGGCGCCTTT GGTGACTTCC TACTCGACAT

1821 AAGAAGCAGA TCGTTCAAAC ATTGGGCAAT AAAGTTCCTT AAGATTGAAT CCTGTGCGCG GTCCTTCCGAT  
 TTTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTTCTAACCTTA GGACAACGGC CAGAACGCTA

1891 GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAATTAACA TGTAAATGCAT GACGTTATTT  
 CTAATAGTAT ATTAAGACA ACTTAATGCA ATTCGTACAT TATTAATTGT ACATTACCTA CTGCAATAA

NotI  
 ~~~~~

FIG. 46D

1961 BssHII  
ATGAGATGGG TTTTATAGAT TAGAGTCCCG CAATATACA TTTAATACGC GATAGAAAAC AAAATATAGC  
TACTCTACCC AAAAATACTA ATCTCAGGGC GTTAAATAGT AAATATAGCG CTATCTTTTG TTTTATATCG

2031 BssHII  
GCGCAACTA GGATAAATTA TCGCGGCGCG TGTCACTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC  
CGCGTTTGGT CCTATTTAAT AGCGGCGGCC ACAGTAGATA CAATGATCTA GCTATTTCGAA GATCTCGGCC

2101 BssHII  
CGGTGGAGT CCAATTCGCC CTATAGTGAG TCGTATTACG CGGCTCACT GGCCGTCGTT TTACAACGTC  
GCCACCTCGA GGTAAAGCGG GATATCACTC AGCATAATGC GCGGAGTGA CCGGCAGCAA AATGTTGCAG

2171 BssHII  
GTGACTGGGA AAACCCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG  
CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA GGGGAAAGC GGTGACCCGC

2241 BssHII  
TAATAGCGAA GAGGCCGCA CCGATCGCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG  
ATTATCGCTT CTCCGGGCGT GGCTAGCGGG AAGGGTTGTC AACCGTCCG ACTTACCCTT TACCCTGCGC

2311 BssHII  
CCCTGTAGCG GCGCATTAAG CGCGGCGGGT GTGTGTGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG  
GGACATCGC CGCGTAAATC GCGCCGCCCA CACCACCAAT GCGCGTCCGA CTGGCGATGT GAACGGTTCG

2381 BssHII  
CCCTAGCGCC CGCTCCTTTC GCTTTCCTCC CGAAGAAAGG GAAGGAAGA GCGGTGCAAG CCGCCGAAAG GGGCAGTTCC  
GGGATCGCGG GGGTCCCTT TAGGGTCCG ATTTAGTGT TTACGGCACC TCGACCCCAA AAAACTTGTAT  
AGATTTAGCC CCCGAGGAA ATCCCAAGCC TAAATCACGA AATGCCGTGG AGCTGGGGTT TTTTGAACATA

2451 BssHII  
TAGGGTGGAG GTTCACGTAG TGGGCCATCG CCCTGATAGA CCGTTTTTTCG CCCTTTGACG TTGGAGTCCA  
ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAGC GGGAAACTGC AACCTCAGGT

XbaI

Clai HindIII

BssHII

FIG. 46E

2591 CGTTCCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCT ACTCGGCTT ATTCCTTTTGA  
 GCAAGAAATT ATCACCTGAG AACAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT

2661 TTTATAAAGG ATTTTGCCGA TTTCCGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG  
 AAATATTCCC TAAAACGGCT AAAGCCGGAT AACCAATTTT TTACTCGACT AAATGTTTTT TAAATTGCGC

2731 AATTTTAA CA AATATTAAC GCTTACAATT TAGTGGCAC TTTTCCGGGA AATGTGCGG GAACCCCTAT  
 TAAAATTTG TTTATAATTG CGAATGTTAA ATCCACCGT AAAAGCCCTT TTACACCGCG CTTGGGGATA

2801 TTGTTTATTT TTCTAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA  
 AACAAATAA AAGATTTATG TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAAGTT

2871 TAATATTGAA AAAGGAAGAG TAGAGTATT CAACATTTCC GTGTCCGCTT TATTCCTTTT TTTGCGGCAT  
 ATPATAACTT TTTCTTCTC ATACTCATAA GTTGTAAAGG CACAGCGGGA ATAAAGGAAA AAACGCCGTA

2941 TTTGCCTTCC TGTTTTGTCT CACCCAGAAA CGCTGGTGA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC  
 AAACGGAAGG ACMAAACGA GTGGTCTTT GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG

3011 ACGAGTGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT  
 TGCTCACCCA ATGTAGCTTG ACCTAGAGTT GTCCCCATTC TAGGAACCTCT CAAAAGCGGG GCTTCTTGCA

3081 TTTCCAATGA TGAGCACTTT TAAAGTCTG CTATGTGGG CCGTATTATC CCGTATTGAC GCCGGGCAAG  
 AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCATAACTG CGGCCCGTTC

3151 AGCAACTCGG TCGCCGCATA CACTATCTC AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA  
 TCGTTGAGCC AGCGGCGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTTCGT

3221 TCTTACGGAT GGCATGACAG TAAGAGATT ATGCAGTCT GCCATAACCA TGAGTGATAA CACTGCGGCC  
 AGAATGCCTA CCGTACTGTC ATTCCTTTAA TACGTCACGA CCGTATTGGT ACTCACTATT GTGACGCCGG

3291 AACTTACTTC TGACAACGAT CGGAGGACG AAGGAGCTAA CCGTTTTTTT GCACAACATG GGGGATCATG  
 TTGAATGAAG ACTGTTGCTA GCCTCCYGGC TTCTCTGATT GCGGAAAAAA CCGTGTGTAC CCCCTAGTAC

3361 TAACTCGCCT TGATCGTTGG GAACCGGAG TGAATGAAGC CATAACCAAC GACGAGCGTG ACACCACGAT  
 ATGAGCGGA ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGGTGCTA

FIG. 46F



3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAAC TGGGAACTAC TTA CTCTAGC TTCTCCGGCAA  
 CGGACATCGT TACCGTTGTT GCAACGCCCT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAATAG ACTGGATGGA GCGGATPAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT  
 GTTAATATATC TGACCTACCT CCGCCATATTT CAACGTCCTG GTEAAGACGC GAGCCGGGAA GGCAGACCGA

3571 GGTATTATGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTCAGACAC TGGGGCCAGA  
 CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG TAACTGCTG ACCCCGGTCT

3641 TGGTAAAGCCC TCCCGTATCG TAGTTAATTA CACGACGGGG AGTCAGGCAA CTA TGGATGA ACGAAATAGA  
 ACCAATTCGGG AGGGCATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGC TTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCAC TGAAT AAGCATGGT AACTGTGAGA CCAAGTTTAC TCATATATAC  
 GTCTAGCGAC TCTATCCACG GAGTGACTAA TTCGTAAACA TTGACAGTCT GGTTCAAAATG AGTATATATG

3781 TTTAGATTGA TTTAAAAC TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT  
 AAATCTAACT AAATTTTGA GTA AAAATTA AATTTCTTA GATCCACTTC TAGGAAAAC TATTAGAGTA

3851 GACCAAAATC CCTTAAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
 CTGGTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTTTTTCT GCGGTAAATC TGCTGCTTGC AAACAAAAA ACCACCGCTA CCAGCGGTGG  
 AGAACTCTAG GAAAAAAGA CCGGCATTAG ACGACGAACG TTTGTTTTTT TGGTGGCGAT GGTCCGCCACC

3991 TTTGTTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGGA GGTAACTGGC TTCAGCAGAG CGCAGATACC  
 AAACAAACGG CCTAGTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AACTGCTC GCGTCTATGG

4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGCCACCAC TTCAAGAACT CTTAGACACC GCCTACATAC  
 TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCITGA GACATCGTGG CCGATGTATG

4131 CTCGCTGTC TAAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGCTTACC GGGTTGGAAT  
 GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTAC CGCTATTTAG CACAGAAATG CCCAACCTGA

4201 CAAGACGATA GTTACCGGAT AAGGCCAGC GGTGGGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT  
 GTTCTGCTAT CAATGGCCTA TTTCCGGCTC CCAGCCCGAC TTGCCCCCA AGCACGTG TCGGGTCGAA

FIG. 46G

4271 GGAGGGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA  
 CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGAATGCGCA CTCGATATC TTTCCGGGTG CGAAGGGGCTT

4341 GGGAGAAAGG CGGACAGGTA TCCGFPARG GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG  
 CCCTCTTTCC GCCTGTCCAT AGGCCATFCG CCGTCCAGC CTTGTCTCTT CCGGTGCTCC CTCGAAGGTC

4411 GGGGAAACGC CTGGTATCTT TATATCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG  
 CCCCTTTGCG GACCATAGAA ATATCAGGAC AGCCAAAGC GGTGGAGACT GAACTGCGAG CTAAAAACAC

4481 ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGGGCCCT TTTTACGGTT CCTGGCCCTTT  
 TACGAGCAGT CCCCCCGCCT CCGATACCTT TTTCCGGTCC TTGCGCCGGA AAATGCCAA GGACCCGGAAA

4551 TGCTGGCCTT TTGCTCACAT GTTCTTTCTT GCGTTATCCC CTGATTTCTGT GGATARCCGT ATTACCCGCT  
 ACGACCGGAA AACGAGTGT CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

4621 TTGAGTGAGC TGATACCCGT CGCCGCGACC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA  
 AACTCACTCG ACTATGCGA GCGGCGTCCG CTTGCTGGCT CGCCTCGCTC AGTCACTCGC TCCITTCGCCCT

4691 AGAGCGCCCA ATACGCAAC CGCCTCTCCC CGCGGTTGG CCGATTCANT AATGCAGCTG GCACGACAGG  
 TCTCGCGGGT TATGCGTTG GCGGAGAGGG GCGCGCAACC GGCYAAATYAA TTACGTCGAC CGTGTGTCTC

4761 TTTCCCGACT GGAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC  
 AAAGGGCTGA CCTTTCCGCC GTCACCTCGG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGGG

4831 AGGCTTTACA CTTTATGCTT CCGCTCGTA TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACACAGG  
 TCCGAAATGT GAAATACGAA GCGCGAGCAT ACNACACACC TTAACACTCG CCTATTGTTA AAGTGTCTCC

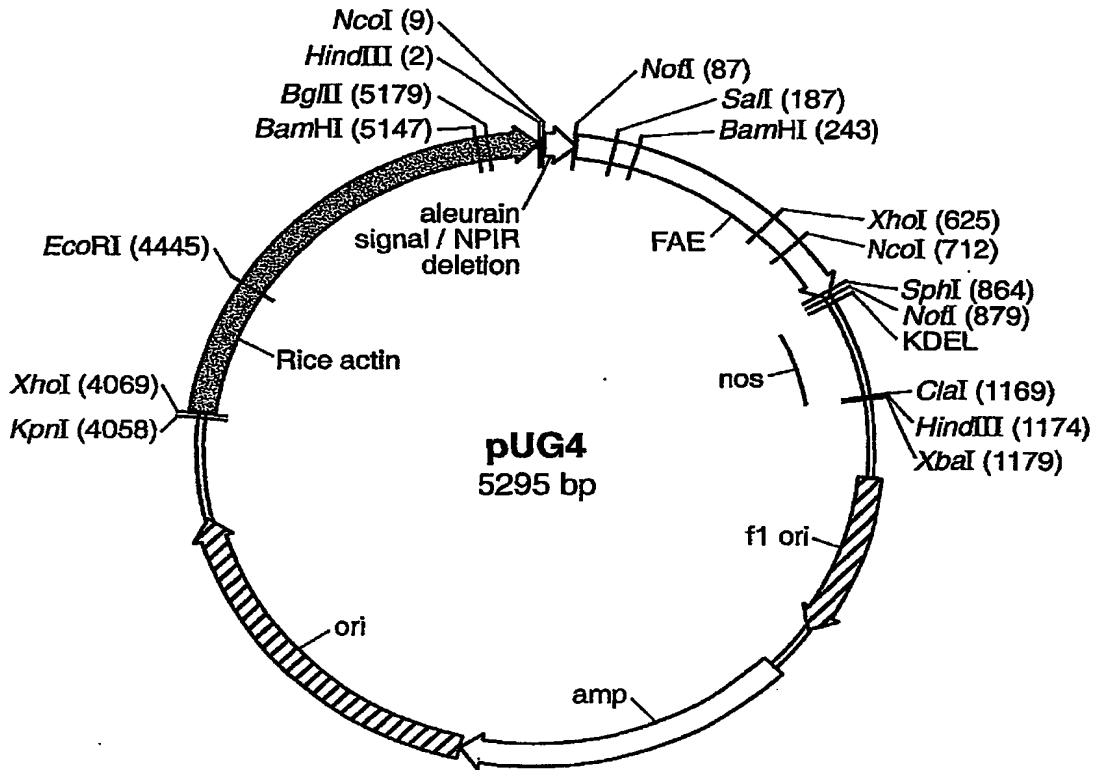
4901 AAACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAAC CCTCACATAA GGGAACAAA GCTGG  
 TTTGTGATA CTGGTACTAA TCGGTTCCG CGGTAAATG GGAGTGATTT CCGTTGTTTT CGACC

EcoRI

BssHII

\*\*\*\*\*

FIG. 46H



**FIG. 47A**

NgOI

HindIII

M A H A R V L L L A L A V L A T A A V A V  
1 AAGCTTACCA TGGCCACGC CCGGTCCTC CTCCTGGCGC TCGCCGTGCT GGCACAGCC CCGGTGCGCC

NotI

A S R A A S T Q G I S E D L Y S R L V E M  
71 TCGCCTCCTC CCGCGCGCC GCGTCCAGC AGGCATCTC CGAGACCTC TACAGCCGTT TAGTCGAAT

Sali

BamHI

A T I S Q A A Y A D L C N I P S T I I K G E K  
141 GGCCACTATC TCCCAAGCTG CCTACGCCA CCGTGTCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA

BamHI

I Y N S Q T D I N G W I L R D D S S K E I I T V  
211 ATTACAAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAAATCACC

F R G T G S D T N L Q L D T N Y T L T P F D T  
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTCGA TACTAACTAC ACCCTCACGC CTTTCGACAC

L P Q C N G C E V H G Y Y I G W V S V Q D Q  
351 CCTACCACAA TGCACCGGTT GTGAATACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA

V E S L V K Q Q V S Q Y P D Y A L T V T G E X L  
421 GTCGAGTCGC TTGTCAAACA GCAGGTAGC CAGTATCCGG ACTACGCGGT GACCGTGACC GGCCACKCCC

G A S L A A L T A A Q L S A T Y D N I R L Y T  
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TGGGACATAC GACAAACATCC GCCTGTACAC

XhoI

F G E P R S G N Q A F A S Y M N D A F Q A S S  
561 CTTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTTCGGTCCG TACATGAACG ATGCTTCCA AGCCTCGAGC

P D T T Q Y F R V T H A N D G I P N L P P V E Q  
631 CCAGATACGA CGCAGTATT CCGGTCACT CATGCCAAGC ACGGCATCCC AAACCTGCC CCGGTGGAGC

NgOI

XhoI

FIG. 47B

. G Y A H G G V E Y W S V D P Y S A Q N T F V C .  
 701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACCA CATTGGTCTG  
 . T G D E V Q C C E A Q G G Q G V N A H T T Y  
 771 CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT

NotI

SphI

F G M T S G A C T W P V A A A E P L K D E L \*  
 841 TTTGGGATGA CGAGCGCGC ATGCACCTGG CCGTTCGCGG CCGCGGAACC ACTGAAGGAT GAGTGTAAA  
 911 GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTCTTTAA GATTGAATCC TGTTCGCCGT CTTGCCGATGA  
 981 TTATCATATA APTTCTGTGG AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA CGTTAATTAT  
 1051 GAGATGGGTT TTATATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAAACAA AATATAGCGC

HindIII

Clal

XbaI

1121 GCAAACTAGG ATAAATATC GCGCGGGTG TCACTAAGT TACTAGATCG ATAAGCTTCT AGAGGGGCGG  
 1191 GTGGAGCTCC AATTGCCCCT ATAGTGAGTC GTATFACCGG GTATFACCGG CCGTCACATGG CCGTTCGTTT ACAACGTCGT  
 1261 GACTGGGAA ACCCTGGCGT TACCCTAACC AATCGCCTG CAGCACATCC CCGTTCGCC AGCTGGCGTA  
 1331 ATAGCGAAG GCGCCGACC GATCGCCCTT CCCAACAGT GCGCAGCCTG AATGGCGAAT GGGACCGCGC  
 1401 CTGTAGCGGC GCATTAAGCG CCGCGGGTGT GGTGGTACG CCGTACACT CGGCTACACT TCGCAGCGCC  
 1471 CTAGCGCCG CTCCCTTTCCG TTTCTTTCCCT TCCTTTCTCG CCACGTTCCG CCGCTTTCCG CGTCAAGCTC  
 1541 TAAATCGGG GCTCCCTTTA GGGTTCCGAT TTACTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA  
 1611 GGGTGATGGT TCACGTACTT GTTCCAAACT GGTGACAGC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG  
 1751 TTCTTTAATA GTGGCCGATT TCGGCCATTT GGTGAAAGT TGAGCTGATT TAACAAAAT TTAACGCGAA  
 1821 TATAAAGGAT TTTGCCGATT TACAAATTTA GGTGCACTT TTCGGGAAA TGTGCGCGGA ACCCTATTTT  
 1891 GTTATTTT CTAATACAT TCAATATGT ATCCGTCAT GAGACAATA CCCTGATAAA TCGTTCATA  
 1961 ATATTGAAA AGGAAGAGTA TGAATTTCA ACAATTCCTG GTCGCCCTTA TTCCCTTTT TCGGCCATTT  
 2031 TGCCTTCTG TTTTGTCTCA CCCAGAAACG CTGGTAAAG TAAAGATGC TGAAGATCAG TTGGGTGCAC  
 2101 GAGTGGGTTA CATCGAATG GATCTCAACA CCGTAAAGT CCTTGAGAGT TTTTCGCCCG AAGAACGTTT  
 2171 TCCAATGATG AGCACTTTTA AAGTTCTGCT ATGTGGCGCG GTATTATCCC GTATTTGACGC CCGGCANGAG  
 2241 CAACTCGGTC CCGCATACA CTATTTCTCAG AATGACTGG TTGAGTACTC ACCAGTACA GAAAAGCATC  
 2311 TTACGGATGG CATTGACAGTA AGAGATTTT GAGCTGTGC CATFACCATG AGTGAFAACA CTGCGGCCAA  
 2381 CTTACTTCTG ACAACGATCG GAGGACCGAA GGAAGTTTACC GCTTTTTCG ACAACATGGG GATCATGTA  
 2451 ACTCGCCTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCTGTGAC ACCACGATGC  
 2521 CTGTAGCAAT GGCAACAACG TTGGCGCAAC TATTAACCTG CGAACTACTT ACTCTAGCTT CCCGGCAACA

FIG.-47C

2591 APTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTCTCTGGCT CGGCCCTTCC GGTGGCTGG  
2661 TTTATTTGCTG ATAAATCTGG AGCCGGTGG CFTGGTCTC CCGGTATCAT TGCAGGACTG GGGCCAGATG  
2731 GTAAGCCCTC CCGTATCGTA GTTATCTACA CCAAGGGGAG TCAGGCAACT ATGGATGAC AATATAGACA  
2801 GATCGCTGAG ATAGGTGCCCT CACTGATTA GCATTGGTAA CTGTTCAGACC AAGTTTACTC ATATATACTTT  
2871 TAGATTGAT TAAACTTCA TTTTAAATTT AAAAGGATCT AGGTGAAGT CCTTTTNGAT AATCTCATGA  
2941 CCAAAATCCC TTAACGTGAG TTTTCGTTC ACTTGAGCTC AGACCCCGTA GAAAGATCA AAGGATCTTC  
3011 TTGAGATCCG TTTTTCCTGC GCGTAATCTG ACCAATCTTT TTTCCGAAG CAACGATCC CAGCGTGGTT  
3081 TGTTCGCCG ATCAAGACT TCTAGTGTAG CCGTAGTTT GCGTACTCT TAAGAATCT GTAGCCCGC CAGATACCAA  
3151 ATACTGTCTT TCTAGTGTAG CAGTGGTGC TGCAGTGGC GATAAGTCT GATTTACCGG GTTGGACTCA  
3221 CGCTCTGCTA ATCTGTACT TACCGGATA GCGCAGCGG TCGGGCTGAA CCGGGGGTTC GTGCACACAG CCCAGCTTGG  
3291 AGACGATAGT TACCGGATA TACCAGTACC CTGAGATGTA TACAGCTGA GCTATGAGAA ABGCCACGC TTCCCGAAGG  
3361 AGCGAACGAC CTACACCGAA CTGAGATGTA CCGTAAGCGG CAGGTCTGGA ACAGGAGAGC GCACGAGGA GCTTCCAGGG  
3431 GAGAAAGCGG GACAGGTATC GGTATCTTTA TAGTCTGTG GGTTCCTGC CACTCTGACT TGAGCCCTCGA TTTTGTGTAT  
3501 GAAACGCCCT GGTATCTTTA TAGTCTGTG GGTTCCTGC CACTCTGACT TGAGCCCTCGA TTTTGTGTAT  
3571 GCTCGTCAGG GGGCGGAGC CTATGGAAA ACGCCAGCA CGCGGCCCTT TTACGGTTC TGGCCCTTTG  
3641 CTGGCTTTT GCTCACATGT TCTTCTCTGC GTFATCCCCT GATTCGTGG GATTCGTGG TACCCTCTTT  
3711 GAGTGAGCTG ATACCGCTCG CCGCAGCGG ACAGCCGAGC GCAGCGAGT AGTGAGCGAG GAAGCGGAAG  
3781 AGCGCCCAAT ACAGCCGCTG CCTCTCCCG CCGGTTCGCC GATTCATTA TGCAGCTGGC ACAGACAGTT  
3851 TCCCGACTGG AAAGCGGCA GTGAGCGCAA CGCAATTA GTGAGTTAG TCACTCATTA GGCACCCCG  
3921 GCTTTACTT TTAGTCTTCC GCTCTGATG TTGTGTGGA TTGTGTGCGG ATACAATTT CACACAGGAA  
  
3991 ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCCG  
  
4061 GCGCCCTC GAGGTCTC ATATGCTTGA GAAGAGATC GGGATAGTCC AATAAANAAC AAAGGTAAGA  
4131 TACCTGGTC AAAAGTGAA ACATCAGTTA AAAGTGGTA TAAGTAAAT ATCGGTAATA AAAGGTGGCC  
4201 CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAAATGA GGATGTTTTG TCGGTACTTT GATACGTCTAT  
4271 TTTTGTATGA ATTTGGTTTT AAGTTTATC GCGATTTGGA AATGCATATC TGTATTTGAG TCGGTTTTTA  
4341 AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT  
  
4411 GACATAAATTT TTGAGAAAA TATAATTTA GCGGAATTC ACAATGAACA AATAAAGAT TAAAAATAGCT  
4481 TGCCCCCGTT GCAGCGATGG GTATTTTTTC TAGTAAATA AAAGTAAAC TTAGACTCAA AACATTTACA  
4551 AAAACAACCC CTAAGTCTT AAAGCCCAA GTGCTATGCA CGATCCATAG CAGCCCATG CCAACCCAAC  
4621 CCAACCCAAC CCACCCCACT GCAGCCCACT GCGAAATAGT CTCCACCCCC GGCACACTCA CCGTGTAGTTG

KpnI

XhoI

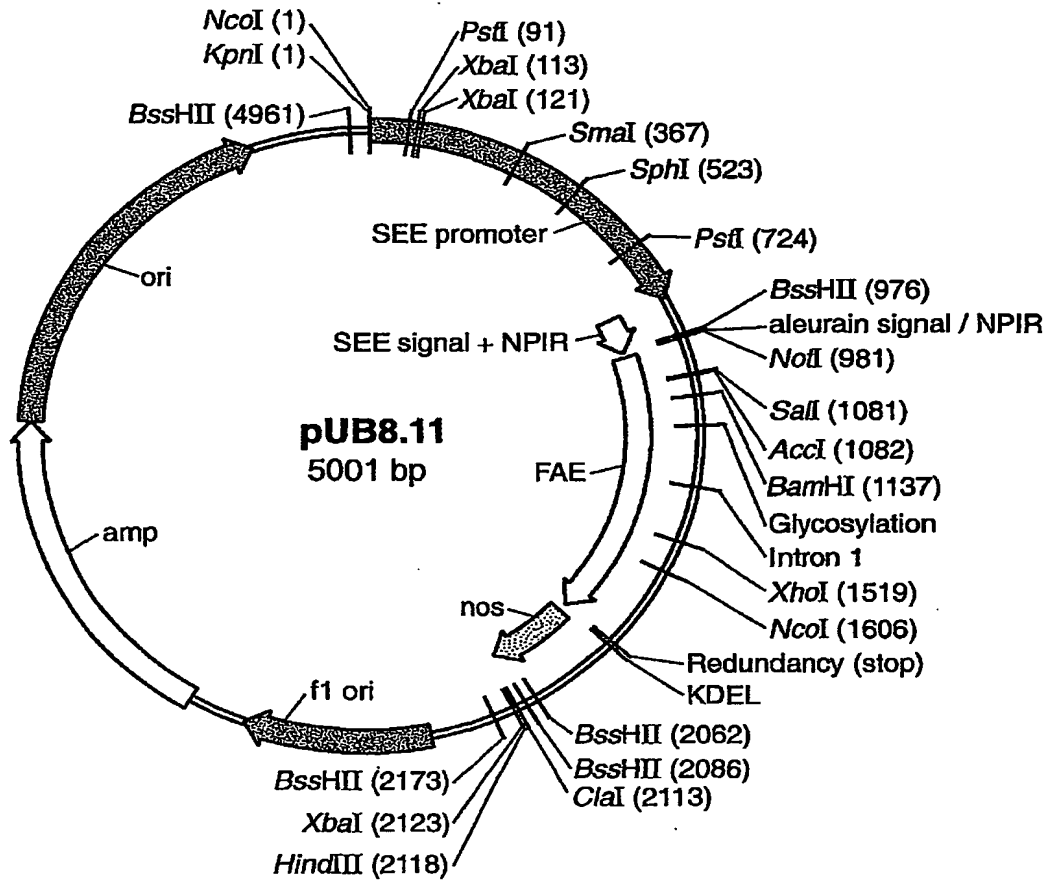
EcoRI

FIG.-47D

... (R) III F 26)

4691 TCCGCACCAC CGCACGTC TC GCAGCCAAA AAAAAAAG AAGAAAAAAGA AAAACAGCAG  
 4761 GTGGTCCGG GTCGTGGGG CCGAARAAG GAGGAGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC  
 4831 CGCTTCCAAA GAAACGCCCC CCAFCGCCAC TATAACATA CCCCCCTC TCTCCCATC CCCCCAACCC  
 4901 TACCACCACC ACCACCACCA CCTCCTCCC CCTCGCTGCC GGACGACGAG CTCTCCCCC CTCCCCCTCC  
 4971 GCCGCCGCC GTAACCAACC CGCCCTCTC CTCTTCTTT CTCCGTTTT TTTTTGCTT CGGTCTCGAT  
 5041 CTTTGGCCTT GGTAGTTGG GTGGCGAGA GCGGCTCTGT CGCCACAGATC GGTGCGCGGG AGGGGCGGGA  
BamHI  
~~~~~  
 5111 TCTCGCGGCT GCGGTCCTCC GCGCTGAGTC GGCCCGGATC CTCGGCGGGA ATGGGGCTCT CGGATGTAGA  
BglII  
~~~~~  
 5181 TCTTCTTCT TTCTTCTTTT TGNGGTAGAA TTGAAATCCC TCAGCATTTG TCAATCGGTAG TTTTTCTTTT  
 5251 CATGATTTGT GACAAATGCA GCCTCGTSCG GAGCTTTTT GTAGC

FIG.-47E



**FIG. 48A**



```

NcoI
*****
KpnI
~
1  CATGGGCCAG  GTATATATAT  GGGATATCTC  AAGCAATAAA  TCGAAATATC  ACCATTGGCT  ACAATATCTG
   GTACCCGGTC  CATATTATAA  CCTATATAG  TTCGTTTATT  AGCTTATAG  TGGTAACCGA  TGTATATAGC

      PstI
      *****
      XbaI
      *****
      XbaI
      *****
71  AGCTCCGAGT  TCTGACTGCA  GTC TGGATGA  CGCGTGTGT  ATCTAGAACT  CTAGATAGCA  CAGCCACAGC
   TCGAGGCTCA  AGACTGACGT  CAGACCTACT  GCGCACAACA  TAGATCTTGA  GATCTATCGT  GTCGGGTGTC

141  ACCTACAGGA  GTGCGACACT  TGTGGACTGT  AGTAGTGTG  GAGACGGAGC  TCTTTCCTAC  CTCCTGACGT
   TGGATGTCC  CACGCTGTGA  ACACCTGACA  TCATCACAA  CTTCTGCTCG  AGAAAGGATG  GAGGACTGCA

211  TCGCCCGGTT  GTCCATYCCA  ACGGCATCAC  TCTCAACCAA  TCACGGCTC  CCAACAAAAT  ATCGTCCCCC
   ACGGCGGCAA  CAGGTAAGGT  TGCCGTAGTG  AGAGTTGGYT  AGTSCCGGAG  GGTGTGTTTTA  TAGCAGGGGG

281  ATGTCCTGGC  GGAGAGAGAG  TACATACATG  CTGTGCGGCC  GTTTTGTGCT  GAATCTCGCT  TCCACTGGCC
   TACAGAACCG  CCTCTCTCTC  ATGTATGTAC  GACAGCGCGG  CAAAACAGA  CTTAGAGCGA  AGGTGACCGG

      SmaI
      *****
351  AATCAGCTCA  GCTCCCGGGA  GCTCACTCAT  TCAAGATCCC  ATCGTCGTG  TCACCCCTGG  CGTCATGGGA
   TTAGTCCAGT  CGAGGGCCCT  CGAGTGAGTA  AGTTCYAGGG  TAGCAGCAGC  AGTGGGGACC  GCAGTACCCCT

421  TGGAAAAGAA  CCTCCGTTGC  TCGGATGAGT  CAGCCATATC  CCCGAACAGA  GTACTGCAAG  ATAACCCCAAT
   ACCTTTTCTT  GGAGGCAACG  AGCCTACTCA  GTCGGTATAG  GGGCTGTCT  CATGACGTTT  TATTGGGTTA

      SphI
      *****
491  TCAGATTCCC  CCAATAGAGA  AAGTATAGCA  TGCTTTCGGG  TTTTGTGTTG  CTTAATTGAC  TTTATTTTTG
   AGTCTAAGGG  GGTATCTCT  TTCATATCGT  ACGAAAGCCC  AAAACAACC  GAATTTACTG  AAATAAABC

561  TTGGAGTTGA  ATGCTGATTT  GTTGTGTA  AATGCCAAC  ATCTGAATAT  CGAGACGGAT  AATAGGCTGG
   AACCTCAACT  TAGGACTAAA  CAACACATTT  TACGGGTTGG  TAGACTTATA  GCTCTGCC  TTTATCCGACC

```

FIG. 48B

631 CTAATTAAT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGCATTACA GCTGGAGCCT  
 GATTAATTA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI

\*\*\*\*\*

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGARGCGTGG CGATGAGATG GGTATATAAAC  
 AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCTGTTACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771 CCCCCGCACC GGGACGGGAG CTCCCBCCTA CCAGTACCAT CTGGCCTCGC TCCCCCTGCC GGACGACCCA  
 GGGGCCGTGG CCCTGGCTC GAGGGCGGAT GGTCAATGGA GAGCGGAGCG AGGGGACCGG CCTGCTGGGT

841 GTAATAACT GTTCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT  
 CATTTTATGA CAACGGTGA CGGCCGCTC TACCGGTGC CGCGTAGGA GAAGAACC GC GAGCGGACGA

B88HII

\*\*\*\*\*

NotI

\*\*\*

911 TGGCCACCGC CGGGTGGC GCCGCATCNT TGGGGACTC CAACCCGATG CGGCCCGTCA CCGAGCGCGC  
 ACCGGTGGC GCGCCACCG CGCGTAGNA ACCGCCTAG GTTGGCTAG GCCGGCAGT GGCTCGCGCG

NotI

\*\*\*\*\*

981 GGCCGCTCC ACGCAGGCA TCTCCGAAGA CCTCTACAGC CGTTAGTGC AATGGCCAC TATCTCCAA  
 CCGCGGAGG TCCCTCCCGT AGAGCTTCT GGAGATGTC GCAAAATCAGC TTTACCCGGT ATAGAGGTT

SaI

\*\*\*\*\*

AccI

\*\*\*\*\*

1051 GCTGCCYACG CCGACCTGTG CAACATTCG TCGACTATA TCAAGGGAGA GAAAATTAC AATTCACAA  
 CGACGGATGC GGCTGGACAC GTTGTAGGC AGCTGATAAT AGTTCCTCT CTTTAAATG TTAGAGTTT

FIG. 48C

BamHI  
\*\*\*\*\*

1121 CTGACATTA CCGATGGATC CTCGGGACG ACAGCAGCAA AGAATAATC ACCGTCCTCC GTGGCACTGG  
 GACTGTAATT GCCTACCTAG GAGCGCTGC TGCCTCGTT TCTTTATTAG TGGCAGAAGG CACCGTGACC

1191 TAGTGATACG AATCTACAAC TCGATACTAA CTACACCCTC AGCCCTTTCG ACACCCCTACC ACAATGCAAC  
 ATCACTATGC TTAGATGTTG AGCTATGATT GATGTGGAG TCGCGAAAGC TGTGGGATGG TGTTFACGTTG

1261 GGTGTGGAAG TACACGGTGG ATATTATATT GGATGGGCTT CCGTCCAGGA CCAAGTCCGAG TCGCTTGTCA  
 CCAACACTTC ATGTGCCACC TATAATATAA CCTACCAGA GGCAGGTCCT GGTTCAGCTC AGCGAACAGT

1331 AACAGCAGGT TAGCCAGTAT CCGGACTACG CGTGTGACCGT GACCGGCCAC KCCCTCGGC CCTCCCTGGC  
 TTGTCTGCCA ATCGGTGATA GGCCTGATGC GCGACTGGCA CTGGCCGGTG MGGGAGCCGC GGAGGGACCG

1401 GGCACCTCACT GCCGCCCAGC TGTCTGGGAC ATACGACAAC ATCCGCCCTGT ACACCTTCGG CGAACCCTGGC  
 CCGTGAATGA CCGCGGGTCC ACAGACGCTG TATGCTGTTG TAGCGGACA TGTGGAAAGC GCTTGGCCGG

1471 AGCGGCATC AGGCCCTCGC GTCGTACATG AACGATGCTT TCCAAGCCTC GAGCCCAGAT ACGACGCAAT  
 TCGCCCGTTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGAG CTCGGGTCTA TGCCTGCTCA

XhoI  
\*\*\*\*\*

1541 ATTTCCGGGT CACTCATGCC AACGACGGCA TCCCAAACTT GCCCCCGGTG GAGCAGGGGT ACGCCCATGG  
 TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGGTTTGGG CCGGGGCCAC CTCGTCCCCA TCGGGTACC

1611 CGGTGTAGAG TACTGGAGCG TTGATCCTTA CAGCGCCAG AACACATTTG TCTGCACCTGG GGAATGAAGTG  
 GCCACATCTC ATGACCTCGC AACTAGGAAT GTCCGGGGTC TTGTGTAAC AGACGTGACC CCTACTTTCAC

1681 CAGTGTGTG AGGCCAGGG CGGACAGGT GTGAATAATG CGCACAGGAC TTATTTTGGG ATGACGAGCG  
 GTCACGACAC TCCGGGTCC GCCTGTCCA CACTTATTTAC GCGTGTGCTG AATAAAACC TACTGTCTGC

1751 GAGCCTGTAC ATGGTGAATCA GTCATTTTCAG CCTCCCCGAG TGTACCAGGA AAGATGGATG TCCTGGAGAG  
 CTCGGACATG TACCCTAGT CAGTAAAGTC GGAGGGGCTC ACATGGTCTT TTCTACCTAC AGGACCTCTC

NcoI  
\*\*\*\*\*

FIG. 48D

1821 GGGGCCGGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTTGGCA ATAAAAGTTTC  
 CCCCAGCGCA TTGGTACTT CCTACTCGAC AFTTCTTCGT CTAGCAAGTT TGTAAACCGT TATTTCAAAG

1891 TTAAGATTGA ATCCTGTTGC CGGTCTTGGC GGCAGAACGC TACTAATAGT ATATTAAAGA CAACTTAAATG CAATTCGGTAC  
 AATTCCTAAT TAGGACAACG GCCAGAACGC TACTAATAGT ATATTAAAGA CAACTTAAATG CAATTCGGTAC

1961 TAAATAATTAA CATGTAATGC ATGACGTTAT TTAGAGATG GGTTTTTATG ATTAGAGTCC CGCAATTATA  
 ATTATTAAAT GTACATTACG TACTGCAATA AATACTCTAC CCAAAAATAC TAATCTCAGG CGGTTAATAT

2031 CATTAAATAC GCGATAGAA ACAAATATA GCGCGCAAC TAGGATAAAT TATCGCGCGC GGTGTCAATCT  
 GTAAATATG CGCTATCTTT TGTTTTATAT CGCGCGTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

XbaI  
 \*\*\*\*\*

ClaI HindIII  
 \*\*\*\*\*

2101 ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGTTGGAG CTCCAATTCG CCTATATAGT AGTCGTATTA  
 TACAATGATC TAGCTATTCG AAGATCTCGC CGCCACCCTC GAGTTAAGC GGGATATCAC TCAGCATAAAT

BssHII  
 \*\*\*\*\*

2171 CCGCGGCTCA CTGGCCGTCG TTTTACAACG TCGTACTGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC  
 GCGCGCGAGT GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTGTGGAC CGCAATGGGT TGAATTAGCG

2241 CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAC  
 GAACGTCGTG TAGGGGGAAA GCGGTCGACC GCATTAATCGC TTCTCCGGC GTGGCTAGCG GGAAGGGTTG

2311 AGTTGCGCAG CCTGAATGGC GAATGGGACG CGCCCTGTAG CGGGCATTG AGCGGGCGCG GTGTGGTGGT  
 TCAACGGGTC GGAATTACCG CTTACCCTGC GCGGACATC GCCGCGTAT TCGGCGCCGC CACACCACCA

2381 TAGGGCAGC GTBACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCTTTCCTTT  
 ATGCGCGTCG CACTGGCGAT GTGAACGGTC GCGGATCGC GGGCGAGAA AGCGAAAGAA GGGAAAGAAA

FIG. 48E

2451 CTCGCCACGT TCGCCGGCCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTTAGTG  
 GAGCGGTGCA AGCGGCCGAA AGGGCCAGTT CGAGATTTAG CCCCCGAGGG AAATCCCAAG GCTAAATCAC

2521 CTTTACGGCA CCTCGACCC AAAAAACTTG ATTAGSGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA  
 GAAATGCCGT GGAGCTGGGG TTTTGTGAAC TAAATCCCACT ACCAAGTGA TCACCCGGTA GCGGGACTAT

2591 GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CAGTCTCTTT AATAGTGGAC TCTTGTTCGA AACTGGACA  
 CTGCCAAAA GCGGGAACCT GCNACCTCAG GTGCAAGAAA TTATCACCTG AGAACAAAGT TTGACCTTGT

2661 ACACCTCAAC CTATCTCGGT CTATCTCTTT GATTTATAAG GGATTTGCC GATTTCCGGC TATTTGGTTAA  
 TGTGAGTTGG GATAGAGCCA GATAAGAAA CTAATATTC CTTAAAACGG CTAAGCCGG ATAACCAAT

2731 AAAATGAGCT GATTTAACAA AAATTTAACC CGAATTTTAA CAAAATATTA ACGCTTACA TTTTAGTGGC  
 TTTTACTCGA CTAATTTGTT TTTAAATGTC GCTTAAAAAT GTTTTATAAT TCGGAATGTT AAATCCACCC

2801 ACTTTTCCGG GAATGTCCG CGGAACCCCT ATTTGTTTAT TTTCTFAAT ACATTCRAAT ATGTATCCGC  
 TGAAAAGCCC CTTTACACGC GCCTTGGGA TAAACAATA AAAAGATTTA TGTAAAGTTTA TACATAGGCG

2871 TCATGAGACA AFAACCTGA TAAATGCTTC AATAATATG AAAAAGGAAG AGTATGAGTA TTCAACATTT  
 AGTACTCTGT TATTTGGACT ATTTACGAAG TTATTATAAC TTTTCTCTTC TCATCTCAT AAGTTGTAAA

2941 CCGTGTCCG CTTATTCCT TTTTTCGGC ATTTTGCCTT CCTGTTTTG CTCACCCAGA AACGCTGGTG  
 GGCACAGCG GAATAAGGGA AAAAACGCCG TAAACGGAA GGACAAAAAC GAGTGGGTCT TTGGGACCAC

3011 AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
 TTTTCATTTT TACGACTTCT AGTCAACCCA CGTGTCTACC CAATGTAGCT TGACCTAGAG TTGTGCCCCAT

3081 AGATCCTTGA GATTTTCGC CCCGAAGAAC GTTTTCCAAAT GATGAGCACT TTTAAAGTTC TGTATGTGG  
 TCTAGGAAC CTCAAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG ACGATACACC

3151 CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCCCGCA TACACTATTC TCAGAAATGAC  
 GCGCCATAAT AGGCATPAAC TGCGGCCCGT TCTCGTTGAG CCAGCGCGGT ATGTGATNAG AGTCTTACTG

3221 TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACCG ATGGCATGAC AGTAGAGAA TTATGCAGTG  
 AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAAATGCC TACCGTACTG TCAATCTCTT AATACATCAC

FIG.-48F

3291 CTGCCATAAC CATGAGTGAAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT  
 GACGGTATTG GTACTCACTA TTGTGACGCC GGTGGAATGA AGACTGTTCG TAGCCTCCTG GCTTCCCTCGA

3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACCTCG CTTGATCCTT GGGAAACCGA GCTGAANTGA  
 TTGGCGAANA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAACHTAGCAA CCCTTGGCCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCTGTAG CAATGGCAAC AACGTTGCGC AAACHTATTA  
 CCGTATGGTT TGCYGTCCG ACTGTGGTGC TACGGACATC GTTACCCGTTG TTGCAACCGG TTTGATAATT

3501 CTGGCGAACT ACTTACTCTA GCTTCCCGG AACAATTAAT AGACTGATG GAGGCGGATA AAGTTGCAGG  
 GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCCTAT TTCAACGCTCC

3571 ACCACTTCG CGCTCGGCC TTCCGGCTGG CTGGTTTAT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG  
 TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAATTA CGACTATTTA GACCTCGGCC ACTCGCACCC

3641 TCTCGCGGTA TCATTCGAGC ACTGGGGCCA GATGGTAAG CCTCCCGTAT CGTAGTTATC TACACGACGG  
 AGAGCGCCAT AGTAAACGTCG TGACCCCGGT CTACCATTCC GGAGGGCATA GCATCAATAG ATGTGCTGCC

3711 GGAGTCAGGC AACATATGGAT GAACGAATA GACAGATCG TGAGATAGGT GCCTCACTGA TTAAGCATG  
 CCTCAGTCCG TTGATACCCTA CTTCCTTTAT CTGCTAGCG ACYCTATCCA CGGAGTGAAT AATTCGTAAC

3781 GTAACTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCATTTTTA ATTTAAAAGG  
 CATTGACAGT CTGGTTCAA TGAATATATA TGAATCTAA CTAATTTTG AAGTAAAAAT TAAATTTTCC

3851 ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTTCG TTCCACTGAG  
 TAGATCCACT TCTAGGAAA ACTATATTAG TACTGGTTTT AGGGAATTGC ACTCAAAGC AAGGTGACTC

3921 CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT  
 GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGAAAAAAA GACGGCCTAT AGACGACGAA

3991 GCAACAANA AAACCACCGC TACCAGCGGT GFTTGTGTTG CCGGATCAAG AGCTACCAAC TCTTTTTTCCG  
 CGTTTGTTTT TTTGGTGGCG ATGGTCGCCA CCAACAACA GGCCTAGTTC TCGATGGTTG AGAAAAAGGC

4061 AAGGTRACTG GCTTCAGCAG AGCCAGATA CCAATACTG TCCTTCTTAGT GTAGCCGTAG TTAGGCCACC  
 TTCCATTGAC CGAGTCCGC TCGCGTCTAT GGTATTAGC AGGAAGATCA CATCGGCATC AATCCGGTGG

**FIG.-48G**

4131 ACTTCAAGAA CTCGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG  
 TGAAGTCTTT GAGACATCGT GCGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC GACGACGGTC

4201 TGGCGATAAG TCGTGTCTTA CCGGTTTGA CTCAAGACGA TAGTTACC GG ATAAGGCCA GCGGTCGGGC  
 ACCGCTATTC AGCACAGAAT GGCCCAACCT GAGTTCGTCT ATCAATGGCC TATTCCGCGT CGCCAGCCCG

4271 TGAACGGGGG GTTCGTGCAC ACAGCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCCTACAGC  
 ACTTGGCCCCC CAAACAGTGT TGTGGGTTC AACCTCGCTT GTTGGATGTG GCTTGACTCT ATGGATGTTC

4341 GTGAGCTATG AGAAAGGCC ACCTTCCC AGGGAGAA GCGGACAGG TATCCGGTAA GCGGACGGGT  
 CACTCGATAC TCTTTCGGCG TCGGAAGGC TTCCCTCTTT CCGCTGTCC ATAGGCCATT CGCCGTCCCA

4411 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTATATAGTCC TGTCCGGTTF  
 GCCTTGTCTT CTCGGGTGCT CCTTCAAGG TCCCCCTTTG CCGACCATAG AATATCAGG ACAGCCCAAA

4481 CGCCACCCTCT GACTTGAGCG TCGATTTTG TGAATCTCGT CAGGGGGCG GAGCCTATGG AAAAACGCCA  
 GCGGTGGAGA CTGAACCTGC AGTAAAAAC ACTACGAGCA GTCCCCCGCG CTCGGATACC TTTTGTGGGT

4551 GCAACGGGC CTTTTACGG TTCTGGCCT TTTGCTGGC TTTTGTCTAC AFGTCTTTT CTGCTTATC  
 CGTTGGCGCC GAAAAATGCC AAGAACCGGA AAACGACCGG AAAACGAGTG TACAAGAAAG GACGCAATAG

4621 CCTTGATCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACC GCTGCCGCGAG CCGAACGACC  
 GGGACTAAGA CACTAATGG CATAATGGCG GAACTCACT CGACTATGGC GAGCGGCGTC GGCYTGCTGG

4691 GAGCGCAGCG AGTCAGTGG CGAGGAAGCG GAAGAGCGCC CAATACGCAA ACCGCTCTC CCCGCGCGTT  
 CTCGCGTCCG TCAGTCACTC GCTCCTTCCG CTTCTCGCGG GTTATGGGTT TGGCGGAGAG GGGCGCGCAA

4761 GCGCGATTCA TTAATGCAGC TGGCACGACA GGTTCCCGA CTGGAAGCG GGCAGTGAGC GCAACGCAAT  
 CCGGCTAAGT AATTACCTTC ACCGTGCTGT CCAAAGGGCT GACCTTTCGC CCGTCACTCG CGTTGCGTTA

4831 TAATGTGAGT TAGCTACTC ATTAGGCACC CCAGGCTTTA CACTTATGC TTCCGGCTCG TATGTTGTGT  
 ATTACTCA ATCGAGTGG TAAATCCGTGG GGTCCGAAAT GTGAATATACG AAGGCCGAGC ATACAACA

FIG. 48H

4901 GGAATTGTGA GCGGATAACA ATTTACACACA GGAACACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA  
 CCTPAACACT CGCCTATTGT TAAAGTGTGT CTTTGTGCGA TACTGGTACT AATGGGTTTC GCGCGTTAAT  
 BssHII  
 \*\*\*\*\*

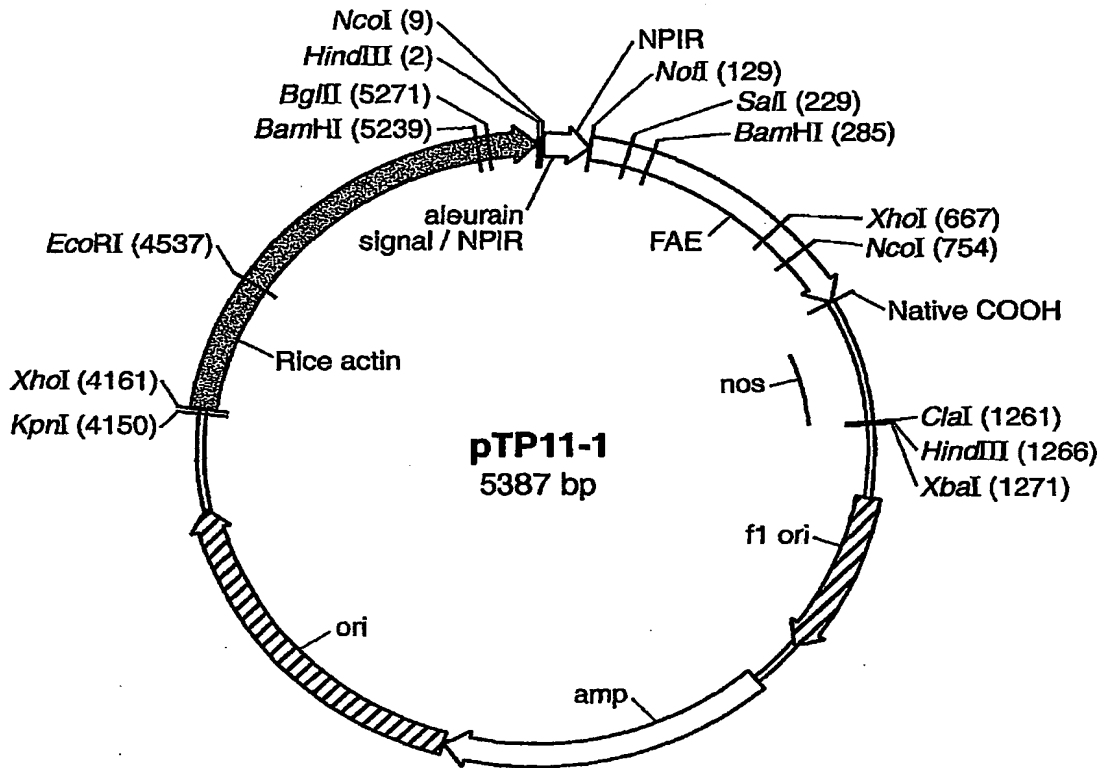
NcoI

KpnI

4971 ACCCTCACTA AAGGGAACA AAGCTGGGTA C  
 TGGGAGTGAT TTCCCTTGTT TTCGACCCAT G  
 \*\*\*\*\*

FIG.\_48I





**FIG. 49A**

NcoI

\*\*\*\*\*

HindIII

\*\*\*\*\*

1 AAGCTTACCA TGGCCACGC CCGCGTCTC CTCCCTGGGC TCGCCGTGCT GGCCAGGCC GCCGTGCGCC  
M A H A R V L L A L A V L A T A A V A V

NotI

\*\*\*\*\*

71 . A S S S F A D S N P I R P V T D R A A A S T .  
TCGCTTCTC CTCCTCTTC GCGACTCA ACCGATCCG GCCGTACC GACCGCGG CCGCTCCAC  
. Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCYCCCAAGC TGCCTACGCC

Sali

\*\*\*\*\*

AccI

\*\*\*\*\*

211 D L C N I P S T I I K G E K I Y N S Q T D I N G  
GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG  
BenHI

\*\*\*\*\*

281 . W I L R D D S S K E I I T V F R G T G S D T N .  
GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCCTGGTA GTGATACGAA  
. L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACAAC TC GATACTAAT ACACCCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGTCTCC GTCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
. Q Y P D Y A L T V T G H X L G A S L A A L T A .  
491 GCCAGTATCC GGACTACGGC CTGACCCGTGA CCGGCCACKC CCTCGGGGCC TCCCTGGCGG CACTCAGTGC  
. A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCTGTAC ACCTTCGGCG AACCGGCGAG CGGCAATCAG

XhoI

\*\*\*\*\*

631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCTCGA GCCAGATAC GACGAGTAT TTCGGGTCA

FIG.- 49B

NcoI  
 \*\*\*\*\*  
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .  
 701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTGA GTGCTGTGAG  
 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGA GCCTGTACAT  
 .  
 911 GGTGATCAGT CATTTCAGCC TCCCAGCTG TACCAGGAAA GATGGATGTC CTGGAGAGGG GGCCCGGTAA  
 981 CCACTGAAGG ATGAGCTGTA AAGAGCAGA TCGTTCAAAC ATTTGGCAAT AAGTTTCTT AAGATTGAAT  
 1051 CCTGTTGCCG GTCCTGCCAT GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAAATTAACA  
 1121 TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG CAATATATACA TTTAATACCG  
 ClaI  
 1191 GATAGAAAC AAAATATAGC GCGCAACTA GGATAAATTA TCGCGCGCGG TGTCACTCTAT GTTACTAGAT

ClaI  
 \*\*\*\*\*  
 XbaI  
 \*\*\*\*\*  
 1261 CGATAAGCTT CTAGAGCGG CCGTGGAGCT CCAATTCGCC CTATAGTGAG TCGTATTTACG CGCGCTCACT  
 1331 GCGCGTCTGT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCACAC TTAATCGCCT TGCAGCACAT  
 1401 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC  
 1471 TGAATGGCGA ATGGGACGCG CCCTGTAGCG CCGCATTAAG CCGCGCGGGT GTGTGGTTA CCGGCAGCGT  
 1541 GACCGCTACA CTTGCCAGCG CCCTAGCGC CCGTCTTTC GCTTCTTTC CTTCTTCTT CCGCCACGTT  
 1611 CCGCGCTTTC CCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG AFTTAGTCT TTACGGCACC  
 1681 TCGACCCCAA AAACTTGTAT TAGGTGATG GTTCACTAG TGGGCCATCG CCCTGATAGA CCGTTTTCG  
 1751 CCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCCT  
 1821 ATCTCGGCTT ATCTTTTGA TTTATAAGGG ATTTGCGCGA TTTTCGGCCTA TTGGTAAAA AATGAGCTGA  
 1891 TTTAACAAAA ATTTAACGG AATTTTACA AAATATTAAC GCTTACAATT TAGTGGCAC TTTTCGGGGA  
 1961 AATGTGCGCG GAACCCCTAT TTGTTTATTTT TTCTAAATAC AFTCAAATAT GTATCCGCTC ATGAGACAAT  
 2031 AACCCGTATA ATGCTTCA TAATATTTAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCCCC  
 2101 TATTCCCTTT TTTGCGGCAT TTTGCTTCC TGTTTTGTCT CACCCAGAAA CGCTGGTGA AGTAAAAGAT  
 2171 GCTGAAGATC AGTTGGGTGC ACGAGTGGGT FACATCGAAC TGGATCTCAA CAGCGTAAAG ATCCTTTGAGA  
 2241 GTTTTCGCCC CGAAGAACGT TTTCCAATGA TTAGTCTCTG TAAAGTTCTG CTAATGTGCG CCGTATTTATC  
 2311 CCGTATTTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTTCTC AGAATGACTT GGTGTGAGTAC  
 2381 TCACCAGTCA CAGAAAAGCA TCCTTACCGAT GGCATGACAG TAAGGAATTT ATGCAGTGTCT GCCATAACCA

FIG. 49C

2451 TGAGTGATAA CACTGCGGCC AACTWACTTC TGACAACGGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT  
2521 GCACAACATG GGGGATCATG TAACTCGCCT TGAATCGTTGG GAACCGGAGC CATACCAAAC  
2591 GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAAC  
2661 TTAATCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGA GCGGATAAA GTTGCAGGAC CACTTCTGCG  
2731 CTCGGCCCTT CCGGCTGGCT GGTATTATTC TGATAAATCT GGAGCCSGTG AGCCTGGTTC TCGCGGTATC  
2801 ATTGAGCAC ATGAGGATGA ACGAANTAGA CAGATCGCTG AGATAGGTGC CTCACGTGAT AACTGTGCAA  
2871 CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAACTTT CATTTTTAAT TTAAGAAGAT CTAGGTGAAG  
3011 ATCTTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG  
3081 TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAAIC TGCTGCTTGC AAACAAAAA  
3151 ACCACCGCTA CCAGCGGTGG TTTGTTTGGC GGATCAAGAG CTACCAACTC TTTTCCGAA GGTAACTGGC  
3221 TTCAGCAGAG CGCAGATACC AATFACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT  
3291 CTGTAGCACC GCCTACATAC CTCGCTCTGC TAACTCTGTT ACCAGTGGCT GCTCGGCTG AACGGGGGT  
3361 GTGTCTTACC GGTTTGACT CAAGACGATA GTTACCAGAT AAGGCGCAGC GGTGCGGCTG AACGGGGGT  
3431 TCGTGCACAC AGCCAGCTT GGAGCGAAG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
3501 AAAGCGCCAC GCTTCCCGAA GGGAGAAAG CCGACAGGTA TCCGGTAAG GGCAGGCTCG GAACAGGAGA  
3571 CCGCACGAGG GAGCTTCCAG ATGCTCGTCA GGGGGCGGGA GCCTATGGAA AAAGCGCAGC AACGGGGCT  
3641 CTTGAGCGTC GATTTTTTGG TGTGCTGCTT TGCTGGCCTT TTGCTCACAT GTTCTTTTCT CTGATTTCTG  
3711 TTTTACCGGT CCGTGGCCTT ATTACCGCT TTGAGTGAGC TGAATCCGCT CGCCGACCGA GCGCAGCGAG  
3781 GGATPAACCGT ATTACCGCT TTGAGTGAGC TGAATCCGCT CGCCGACCGA GCGCAGCGAG  
3851 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAC CCGCTCTCCC CGCGGTTGG CCGATTCATT  
3921 AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAGAGCGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA  
3991 GCTCACTCAT TAGGCACCCC AGGCTTTTACA CTTTATGCTT CCGGCTCGTA TGTGTTGTTG AATTGTTGAGC  
4061 GGATAACAAT TTCACACAGG AACACAGCTAT GACCATGANT ACGCCAAGCG CGCAATTAAC CCTCACTAAA

KpnI  
.....  
XhoI  
.....

4131 GGGAAACAANA GCTGGGTACC GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAAGAGAG TCGGGATAGT  
4201 CCAAAATAAA ACAAAAGGTA GATTACCTGG TCAAAAGTGA AARCAFCAGT TAAAAGGTGG TATAAGTAAA  
4271 ATATCGGTAA TAAAAGGTGG CCCAAAAGTGA AATTTACTCT TTTCTACTAT TATAAATTT GAGGATGTTT  
4341 TGTGCTACT TTAGATACGTC ATTTTGTGAT GAATTTGGTT TTAAGTTTTAT TCGCGATTTG GAAATGCATA  
4411 TCTGTATTG AGTCGGTTTT TAAGTTCCGT GCTTTTGTAA ATACAGAGGG ATTTGHTATAA GAAATATCTT  

EcoRI  
.....

4481 TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAANA AATATATATTT CAGCGAATTT CCACAATGAA  
4551 CAATAAATAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT TCTAGTAAA TAAAAGATAA  
4621 ACTTAGACTC AAAACATTTA CAAAACAAC CCTAAAAGTC CTAAGGCCCA AAGTGTCTATG CACGATCCAT

FIG. 49D

4691 AGCAAGCCCA GCCAAGCCCA ACCAAGCCCA ACCAAGCCCA ACCAAGCCCA ACCAAGCCCA ACCAAGCCCA ACCAAGCCCA  
 4761 CCGGCACATAT CACCGTAGT TGTCCGCACC ACCGACGTC GGTCTGTGGG GGTCTGTGGG GGTCTGTGGG GGTCTGTGGG  
 4831 AAAAGAAAA GAAAACAGC AGGTGGTCC TCCGCTTCCA AAGAAACGCC AAGAAACGCC AAGAAACGCC AAGAAACGCC  
 4901 CGACGAGGCC CGGCCTCCC TCCGCTTCCA AAGAAACGCC CACCTCCTCC CACCTCCTCC CACCTCCTCC CACCTCCTCC  
 4971 TCTCCTTCCA TCCGCTTCCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA  
 5041 AGCTCCTCCC CCAACACCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA CCAACACCA  
 5111 TTTTTTTTCGT CTCGCTTCCG ATCTTTGGCC TTGGTAGTTT GGTGAGCCGA GAGCGGCTTC GTGCGCCAGA  
 BamHI  
 \*\*\*\*\*  
 5181 TCGGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG  
 BglII  
 \*\*\*\*\*  
 5251 GAATGGGCT CTCGGATGA GATCTCTTTT CTTTCTTCTT TTTGGGTAG AATTGAATC CCTCAGCATT  
 5321 GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC

FIG.-49E



771 CCTCCCTCCG CTTCCAAAGA AACGCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC CTCCCATCCC  
GGAGGGAGGC GAAGGTTTCT TTGCGGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCCCTA CCACCACCAC CACCACCACC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCCCCCT  
GGGTTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC CGCCGCCGGT AACCACCCCG CCCCTCTCCT CTTTCTTTCT CCGTTTTTTT TTTCGTCTCG  
GGGGGAGGCG GCGCGGCCA TTGGTGGGGC GGGGAGAGGA GAAAGAAAGA GGCACAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCTG CCCAGATCGG TCGCGGGGAG  
CAGAGCTAGA AACCGGAACC ATCAAACCCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

BamHI

~~~~~

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCGGATCCT CGCGGGGAAT GGGGCTCTCG  
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACCTAGCC GGGCCTAGGA GCGCCCTTA CCCCAGAGC

BglII

~~~~~

1121 GATGTAGATC TTCTTTCTTT CTTCTTTTGG TGGTAGAATT TGAATCCCTC AGCATTTGTC ATOGGTAGTT  
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

~~~~~

1191 TTTCTTTTCA TGATTTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTGTGT AGGTAGAAGC TTACCATGG  
AAAGAAAAGT ACTAAACACT GTTTACGTCTG GAGCACGCCT CGAAAAACA TCCATCTTCTG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,  
pJQ3.2 and pJO6.3.

**FIG. 50B**

**ALEURAIN\_deleted NPIR (Apoplast) Structure and Sequence**



**ALEURAIN-NPIR-DEL**  
93 bp

```

+1      M A H A R V L L L A L A V L A T A A V A
HindIII NcoI
-----
1  AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCGTGCT GCCCAGGCC GCCGTGCGG
   TTCGAATGGT ACCGGGTGCG GGCAGGAG GAGGACCGG AGCGCACGA CCGGTGCCG CCGCAGCGC

+1 V A S S R A A
      NotI
-----
71 TCGCCTCCTC CCGCGCGGCC GCC
   AGCGGAGGAG GCGCGCCCG CGG
    
```

**FIG. 51**



SEE1 ( Senescence enhanced ) PROMOTER sequence

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI                               XbaI   XbaI
      ~~~~~
71  AGCTCCGAGT TCTGACTGCA GTCIGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
      SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCCTTCGGG TTTTGTTTGG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
      PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG

```

**FIG. 52**

SEE1 ( Senescence enhanced ) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI                               XbaI   XbaI
      ~~~~~
71  AGCTCCGAGT TCTGACTGCA GTCIGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
      SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCCTTCGGG TTTTGTTTGG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
      PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
      M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGCCCCAGC GCCGCATCCT CTTCTTGGCG CTCGCCGCTT
      BssHII
      ~~~~~
      NotI
      ~~~~~
911 . A T A A V A A A S L A D S N P I R P V T E R A .
      TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCCG
      NotI
      ~~~~~
981 . A A
      GGCCGCC

```

**FIG. 53**

## INTERNATIONAL SEARCH REPORT

Intern Application No

PCT/US 01/43588

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/82		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) WPI Data, EPO-Internal, PAJ, BIOSIS		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 143 543 A (MICHELSEN BIRGIT ET AL) 7 November 2000 (2000-11-07) see the whole document	1-3, 34-36, 74
Y		4-33, 37-73
Y	DE VRIES R P ET AL: "The faeA genes from Aspergillus niger and Aspergillus tubingensis encode ferulic acid esterases involved in degradation of complex cell wall polysaccharides." APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 63, no. 12, December 1997 (1997-12), pages 4638-4644, XP002203731 ISSN: 0099-2240 see the whole document	4-33, 37-73
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents :		
*A* document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
*E* earlier document but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	
*O* document referring to an oral disclosure, use, exhibition or other means	*&* document member of the same patent family	
*P* document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 27 June 2002	Date of mailing of the international search report 12/07/2002	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer Grosskopf, R	

INTERNATIONAL SEARCH REPORT

Inten Application No

PCT/US 01/43588

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>GARCIA-CONESA MARIA-TERESA ET AL: "A cinnamoyl esterase from Aspergillus niger can break plant cell wall cross-links without release of free diferulic acids." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 266, no. 2, December 1999 (1999-12), pages 644-652, XP002203732 ISSN: 0014-2956</p>	

# INTERNATIONAL SEARCH REPORT

\* Information on patent family members

Intern Application No  
PCT/US 01/43588

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6143543	A	07-11-2000 GB 2301103 A , B	27-11-1996

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT OR DRAWING
- BLURRED OR ILLEGIBLE TEXT OR DRAWING
- SKEWED/SLANTED IMAGES
- COLOR OR BLACK AND WHITE PHOTOGRAPHS
- GRAY SCALE DOCUMENTS
- LINES OR MARKS ON ORIGINAL DOCUMENT
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**